Title
Speakers Communicate Their Perceptual-Motor Experience to Listeners Nonverbally

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 30(30)

ISSN
1069-7977

Authors
Cook, Susan Wagner
Tanenhaus, Micheal K.

Publication Date
2008

Peer reviewed

Speakers Communicate Their Perceptual-Motor Experience to Listeners Nonverbally

Susan Wagner Cook (swcook@bcs.rochester.edu)
Department of Brain and Cognitive Science, University of Rochester
Rochester, NY 14627 USA

Michael K. Tanenhaus (mtan@bcs.rochester.edu)
Department of Brain and Cognitive Science, University of Rochester
Rochester, NY 14627 USA

Abstract

We explored the perceptual motor information expressed by speakers, and encoded by listeners when engaged in naturalistic communication. After solving the Tower of Hanoi task either with real objects or on a computer, speakers explained the Tower of Hanoi task to listeners. Speakers expressed properties of the objects that had been used to solve the task in their hand gestures, but not in their speech. Moreover, listeners were sensitive to speaker's prior experience. In one experiment, listeners who observed explanations from speakers who had previously solved the problem with real objects subsequently treated computer objects more like real objects; their mouse trajectories revealed that they lifted the objects in conjunction with moving them sideways. In a second experiment, listeners were sensitive to the particular constraints imposed by the computer. These findings use the natural behavior of speakers and listeners to provide evidence that both speakers and listeners spontaneously invoke perceptual-motor systems during human communication via spoken language.

Keywords: gesture; communication.

When we reach for objects, the location and shape of our hands reveals information about the location and shape of the object that is the goal of our actions. Surprisingly, something similar happens when we express our thoughts. When people talk, they gesture, changing the location and shape of their hands in precise coordination with what they say. Gestures can represent perceptual (e.g., size or shape) and motor (e.g., movement trajectory) information that is not included in the accompanying speech (Beattie & Shovelton, 1999; McNeill, 1992), including information that is not explicitly part of a speaker's message (Church & Goldin-Meadow, 1986; Goldin-Meadow, 1997). We used hand gesture as a tool for exploring the information expressed by speakers and interpreted by listeners engaged in naturalistic communication.

Human language has been assumed to involve the transmission of abstract and amodal representations (Fodor, 1975; Pylyshyn, 2003; Pylyshyn, 2001). Indeed, an arbitrary relation between form and meaning is often considered a defining feature of linguistic systems (Hockett, 1960, although see Shintel, Okrent & Nusbaum, 2006). In contrast to speech, the gestures produced by speakers in conjunction with speech often appear to express meaning non-arbitrarily, although it is not known: (a) what sorts of representations support production of iconic gesture, and (b) to what degree the resulting gestures are spontaneously encoded by listeners.

We examined both of these issues using the Tower of Hanoi task. Speakers explained the task to listeners, after solving the task either with real objects or on a computer. Speakers' hand gestures, but not their speech, reflected properties of the movements required to move the objects when solving the task, demonstrating that, during language production, gestures can emerge from motor representations. This finding suggests that listeners incorporate some perceptual-motor information conveyed by the speaker's gesture into the representations they construct during spoken language comprehension.

Study 1

Methods

Participants Fourteen pairs of participants were included in the data analysis. Data from an additional two pairs in each condition were eliminated due to irregularities in the experimental procedure.

Procedure Speakers first completed the Tower of Hanoi problem-solving task. In this problem, a stack of disks, arranged from the largest on the bottom to the smallest on top, is arranged on the leftmost of three pegs, and this stack must be moved to the rightmost peg, moving only one disk at a time without placing larger disks on top of smaller disks. Speakers solved the problem using either heavy metal disks on wooden pegs or cartoon pictures on a computer screen. The real objects consisted of four weights (0.6, 1.1, 2.3 and 3.4 kilograms) on a 22.86 x 76.2 cm board with pegs at 12.7, 38.1 and 63.5 cm. The computer objects were presented on a 45.72 cm computer monitor using screen resolution 480 x 640. The real disks needed to be lifted up over the top of the pegs before they could be moved sideways, whereas the computer disks could be dragged horizontally from one peg to another without being lifted over the top of the peg.

After practice solving and explaining three three-disk problems and one four-disk problem to the experimenter, speakers solved the four-disk problem a second time. They then explained how to solve the four-disk problem to the other participant, the listener. This explanation occurred in a different room. After the explanation, listeners solved and explained the four-disk problem twice in the original room. All listeners solved the task on the computer, but speakers were not informed that the listener would be solving the task on the computer. In cases where the speaker had solved the problem using the physical apparatus, a second experimenter surreptitiously hid the apparatus during the explanation. Listeners' mouse trajectories were tracked while they solved the problem (Spivey, Grosjean, & Knoblich, 2005).

Coding Speech and gesture from all participants was recorded, transcribed and coded. We investigated encoding of physical features of the objects in speech across the entire explanation. We coded reference to physical features by identifying all adjectives referring to physical properties of the objects, including color, weight, and physicality. We investigated encoding of physical features in gesture by focusing on participants' description of the movement of the largest disk. We identified the spoken phrases referring to movement of this disk, and subsequently classified the accompanying gestures according to the number of hands used.1 We chose this feature of gesture because physical movement of the largest disk required two hands for speakers who had solved the problem using real objects, but not for speakers who had solved the problem using the computer. Moreover, this feature of gesture could be straightforwardly coded.

Control Study We conducted a control study to independently assess how speakers expressed information about the physical characteristics of the objects used. Eight additional participants were informed about the experimental manipulation, and asked to guess the experimental condition of speakers after listening to the speech without video, or observing the gesture without audio. These participants provided judgments on video-only and audio-only clips of 11 of the participants included in Study 1.

Results

All speakers (N=14) depicted the movement of the disks using hand gestures in conjunction with speech. More importantly, speakers' hand gestures reliably reflected the environment in which they had solved the problem. Speakers who solved the problems with heavy disks generally used two grasping hands when describing movement of the largest disk (5/7 speakers). These speakers could easily have gestured the motion of the largest disk with one moving hand, since they were not actually lifting this disk when gesturing, and they gestured movement of the smaller disks using only one hand. In contrast, speakers who solved the problem on the computer never used two hands when demonstrating movement of the largest disk (0/7 speakers), a pattern reliably different from that seen in speakers interacting with real objects ($\chi^2(1)=7.77$, p=.005, See Figure 2). Thus, speakers' gestures reflected their actual motor experience, rather than a more abstract representation of transfer. In contrast, reference to physical features of the objects in speech was relatively infrequent. Instead, speakers tended to refer to the relative size of the disk in their speech (e.g. "next smallest") regardless of whether the disks were real or presented on the computer.[2]

Figure 1: Examples of single (Computer Condition) and two-handed (Real Objects Condition) gestures referring to the movement of the largest disk.

[1] It is unlikely that this is the only feature of speakers' gesture that distinguished the two groups.

[2] Three participants used the word "heaviest" to refer to the largest disk once. Two of these participants also used the word "lightest" to refer to the smallest disk once. However, use of mass to refer to the computer disks was not infelicitous; one listener also used the word "heavy" when describing their solution to the computer task. Only one participant in the computer condition explicitly referred to the fact that the objects were on the computer screen. Thus, the failure to mention this fact in the real condition was also not infelicitous.

The judgments of our informed raters listening to the audio or viewing the video also suggested that information about the physical features of the objects was available in the gesture and not the speech. We compared the number of correct judgments in each condition with the number that would be expected given chance performance (50%) using a binomial test. On the audio task, participants were not reliably better than chance (48/88 (54%) correct judgments, $p=.45$). In contrast, on the video task, participants were reliably better than chance (54/88 (61%) correct judgments, $p=.04$). Note that the high proportion of errors in the video conditions suggests that the gesture cues were quite subtle, and unlikely to be consciously noted by participants in the primary study, who heard only one explanation, and unlike the control subjects, were unaware of the manipulation.

In order to investigate whether listeners' incorporated motor information that reflected the speaker's experience, we examined the mouse trajectories produced by listeners when they subsequently solved the problem. We analyzed the trajectory of listeners' mouse movements using a maximum likelihood mixed quadratic model to predict the height of a participants' mouse, given the x-coordinate of the mouse movement, the quadratic of the x-coordinate, and the condition under which the speaker had solved the problem. In order to compare height across moves with different starting and ending points, we transformed the x-coordinates on each move so that they ranged from 0 to 1. Data from the first 15 moves produced by each participant were included in the analysis, because at least 15 moves were necessary to solve the problem. Data were analyzed using a mixed model, with y-coordinate as the dependent measure, condition and solution as fixed factors, and subject, x-coordinate, and the intercept as random factors. There was a significant interaction between condition and the quadratic of the x-coordinate ($F(1,10{,}000)=176.21$, $p<.0001$). The function fitting the mouse trajectories had a significantly larger parabolic component for those listeners who had been instructed by speakers who used the real objects in comparison with those listeners who had been instructed by speakers who used the computer (See Figure 3).[3] [4]

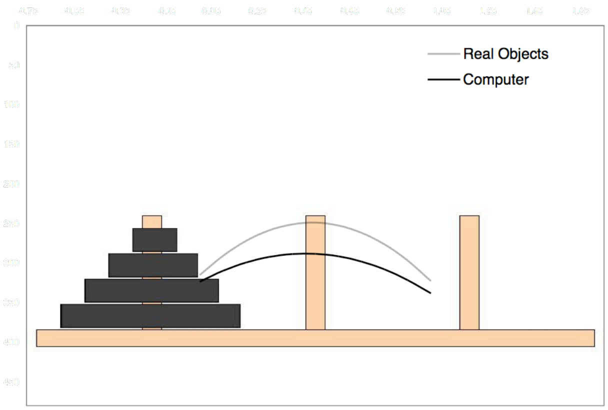

Figure 2: Mouse trajectories predicted using data from the first 15 moves, superimposed on the computer display. The x and y axes are screen coordinates.

Interim Discussion

The findings from Study 1 revealed that speakers' gestures iconicly represent their perceptual-motor experience, even when speakers are not encoding specific perceptual-motor experience in the accompanying speech. Moreover, listeners are sensitive to subtle differences in perceptual motor information encoded in gesture.

One alternative explanation for these findings is that differences in materials, rather than differences in perceptual-motor experience, can account for the results, given that one group used real objects and one group used virtual objects. Study 2 was designed to investigate this possibility while replicating and extending the findings from Study 1.

Study 2

Methods

Participants 24 pairs of individuals participated in this Study. Data from an additional four pairs were eliminated, two because of a data recording error, and two pairs where one participant had participated in prior versions of the study.

Procedure The order of events used in Study 1 was also used in Study 2. However, the nature of the materials used was changed. Speakers again solved the problem using either heavy metal disks on wooden pegs or cartoon pictures on a computer screen. However, there were two groups of speakers who completed the task on the computer. For one group of computer users, the *Computer No Constraints* group, the disks could be moved as in Study 1, in that the disks could be moved sideways without being lifted over the top of the peg on which they were located. In contrast, for the *Computer Constrained* group, the virtual disks exhibited the same sort of constraint as that inherent in the real objects. For these participants, the disks needed to lifted

[3] Additional reliable fixed effects in the model were as follows: There was a reliable quadratic component to the movements ($F(1,12)=2739.03, p<.0001$), a reliable linear component to the movements ($F(1,12)=878.56, p<.0001$), and the linear component interacted with condition ($F(1,10{,}000)=67.99, p<.0001$).

[4] Listeners were equally facile at solving the task across conditions. There was no difference in the number of moves required for the first (Real: 23.9, Comp: 25.1, $t(12)=.25$) or second solution (Real: 39.1, Comp: 26.9, $t(12)=.83$).

above the top of the peg on which they were located before they could be moved sideways.

We also changed the nature of the computer disks the listeners used. All listeners completed the tasks with the constraint that the virtual disks needed to lifted above the top of the peg on which they were located before they could be moved sideways.

Coding Speech and gesture from all participants was recorded, transcribed and coded as in Study 1.

Results

All speakers again depicted the movement of the disks using hand gestures in conjunction with speech. Moreover, the differences in gesture production observed in Study 1 were replicated and extended. Speakers who solved the problems using real objects again used two hands to depict movement of the largest disk (6/7 speakers), while speakers who solved the problem on the computer only used one hand to depict movement of the largest disk (16/17 speakers). Moreover, there also seemed to be a difference in the gestures produced by speakers who had solved the problem on the computer with constraints, in comparison with those speakers who had solved the problem on the computer without constraints. These speakers produced gestures with particularly large, arced trajectories, even when they were not describing movement over the middle peg (see Figure 3).[5]

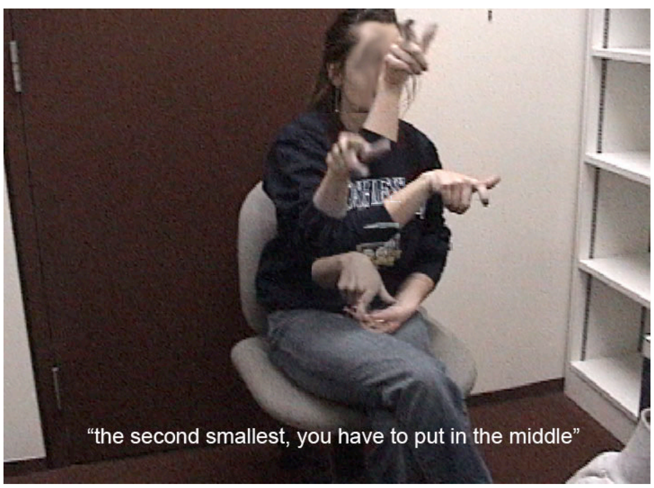

Figure 3: Example of a large, arced gesture produced by a speaker from the Computer Constrained condition.

In order to investigate whether listeners' incorporated speakers' experience, we again examined the mouse trajectories produced by listeners when they subsequently solved the problem. We analyzed the trajectory of listeners' mouse movements using a maximum likelihood mixed quadratic model to predict the height of a participants' mouse, given the x-coordinate of the mouse movement, the quadratic of the x-coordinate, and the condition under which the speaker had solved the problem. Because all listeners were solving the problem with virtual constraints, which required production of arced trajectories, we focused on the initial segment of the first move produced by listeners. In particular, we included all data from the time participants initiated the movement and the time that their mouse was halfway to the second peg. This eliminated reversals in the trajectory that were produced when participants adjusted their movement to accommodate the constraint. Data were analyzed using a mixed model, with y-coordinate as the dependent measure, condition and solution as fixed factors, and subject, x-coordinate, and the intercept as random factors. There was a significant interaction between condition and the quadratic of the x-coordinate ($F(2,495)=18.66$, $p<.0001$). The function fitting the mouse trajectories had a significantly larger parabolic component for those listeners who had been instructed by speakers who used the real objects in comparison with those listeners who had been instructed by speakers in the unconstrained computer condition ($t(495)=6.00$, $p<.0001$), replicating the findings of Study 1. Moreover, the function fitting the mouse trajectories for those listeners who had been instructed by speakers who used the constrained virtual objects were also reliably different from those listeners who had been instructed by speakers who used the unconstrained virtual objects ($t(495)=4.75$, $p<.0001$). (See Figure 3). Thus, listeners' movements were reliably affected by the perceptual and motor experience of the speaker that they observed.

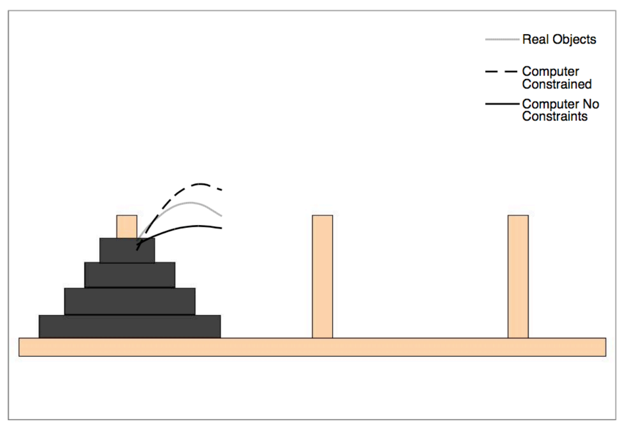

Figure 4: Mouse trajectories predicted using data from the first half of the first move, superimposed on the computer display. The x and y axes are screen coordinates.

As a second window onto listeners' uptake of information, we categorized participants' first move as successful, or unsuccessful, based on whether or not participants needed to reverse trajectory to account for the constraint. Listeners who heard explanations from speakers who had solved the problem on the computer with constraints implicitly adjusted for the constraint in their

[5] Consistent with study 1, we did not find differences in participants' speech across conditions. Across conditions, participants used the same verb to describe the movements, and the same nouns to refer to the blocks.

initial move. Only one (1/9) of these listeners needed to correct their mouse trajectory on their initial move. In contrast, all (8/8) of the listeners who heard explanations from speakers who had solved the problem without constraints needed to correct their trajectory on their initial move, a pattern of performance reliably different from the no constraints group (Fisher's exact test, p=.0001). Listeners who heard explanations from speakers who had solved the problem using the real disks were inconsistent in their initial move (4/7 needed to correct their trajectory).

Discussion

In Study 2, the effect of speakers' prior experience on listeners' subsequent actions was replicated and extended. Listeners reliably moved the disks differently even in the two computer conditions where the materials were held constant. This suggests that the findings from Study 1 are not an artifact of differences in materials across conditions. Instead, it appears that speaker transmit information about the movement affordances of objects to their listeners, and that listeners incorporate this information.

General Discussion

Speakers reflected physical properties of the objects that had been used to solve the task in their hand gestures, revealing activation of perceptual-motor representations in service of communication. Furthermore, this information does not go unheeded, but rather is incorporated into listeners' interpretations. Gestures affected listeners' interpretations, even when participants' gesture was manipulated without direct instruction, was consistent with speech, and was performed and processed spontaneously. This suggests that listeners are also activating perceptual-motor representations when interpreting a speaker's meaning.

These findings are consistent with the hypothesis that gestures may reflect action simulations used by speakers (Hostetter & Alibali, in press). Speakers in the current study who had solved the problem with real objects reliably treated an imaginary disk as if it actually had mass when they communicated about moving that disk. This suggests that speakers were not only activating the goals of their movement, which were expressed in the concurrent speech, but also the specific motor plan that would accomplish these goals, which was expressed in the concurrent gesture.

Listeners in the current study treated virtual disks differently, consistent with the perceptual-motor experience of the speaker that they observed. When interpreting language, even abstract language, people appear to represent perceptual and motor information that is irrelevant to task demands (Glenberg & Kaschak, 2002; Glover, Rosenbaum, Graham, & Dixon, 2004; Kaschak et al., 2005; Richardson, Spivey, Barsalou, & McRae, 2003; Zwaan, Stanfield, & Yaxley, 2002). For example, after hearing a sentence about movement towards their own body, listeners are quicker to make a movement in the same direction (Glenberg & Kaschak, 2002). The activation of perceptual and motor representations in listeners has been inferred from decrements in performance on perceptual-motor secondary tasks. However, these secondary tasks may themselves be reorganizing processing by activating the very systems under investigation. For example, moving one's own body can facilitate recognition of bodily postures in others, by activating aspects of one's own body schema that would otherwise not be recruited (Reed & Farah, 1995). The findings reported here provide evidence that perceptual-motor representations are activated during human communication using data from the natural behavior of speakers and listeners.

These findings suggest that speakers and listeners are not simply communicating abstract and amodal information. Instead, information from the manual modality both reliably reflects speakers' experience in the world and shapes how listeners encode a speaker's message.

Acknowledgments

We thank M. Andrews, N. Cook, M. Hare, K. Housel, A.P. Salverda, and D. Subik for assistance in executing these experiments. This work was supported by NIH grant HD-27206.

References

Arbib, M. A., & Rizzolatti, G. (1996). Neural expectations: A possible evolutionary path from manual skills to language. *Communication & Cognition, 29*(3-4), 393-424.

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., & Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Current Biology, 16*(18), 1818-1823.

Beattie, G., & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language & Social Psychology, 18*(4), 438-462.

Beattie, G., & Shovelton, H. (2001). An experimental investigation of the role of different types of iconic gesture in communication: A semantic feature approach. *Gesture, 1*, 129-149.

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition, 23*(1), 43-71.

Fodor, J. A. (1975). *The language of thought.* Cambridge, MA: Harvard University Press.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*(3), 558-565.

Glover, S., Rosenbaum, D. A., Graham, J., & Dixon, P. (2004). Grasping the meaning of words. *Experimental Brain Research, 154*(1), 103-108.

Goldin-Meadow, S. (1997). When gestures and words speak differently. *Current Directions in Psychological Science, 6*(5), 138-143.

Goldin-Meadow, S., Kim, S., & Singer, M. (1999). What the teacher's hands tell the student's mind about math. *Journal of Educational Psychology, 91*(4), 720-730.

Goldin-Meadow, S., & Singer, M. A. (2003). From children's hands to adults' ears: Gesture's role in the learning process. *Developmental Psychology, 39*(3), 509-520.

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

Hostetter, A. B., & Alibali, M. W. (in press). Visible embodiment: Gestures as Simulated Action. *Psychonomic Bulletin and Review*.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M., Blanchard, A. A., et al. (2005). Perception of motion affects language processing. *Cognition, 94*(3), B79-B89.

Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 28, pp. 389-450). San Diego, CA, US: Academic Press.

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

Pylyshyn, Z. (2003). Return of the mental image: are there really pictures in the brain? *Trends in Cognitive Sciences, 7*(3), 113-118.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*(1-2), 127-158.

Reed, C. L., & Farah, M. J. (1995). The Psychological Reality of the Body Schema - a Test with Normal Participants. *Journal of Experimental Psychology-Human Perception and Performance, 21*(2), 334-343.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science, 27*(5), 767-780.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences, 21*(5), 188-194.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Res Cogn Brain Res, 3*(2), 131-141.

Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech. *Journal of Memory and Language, 55*, 167–177.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences, 102*(29), 10393-10398.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., et al. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience, 17*(2), 273-281.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shape of objects. *Psychological Science*, 13(2), 168-171.