



Published in final edited form as:

Cognition. 2009 October ; 113(1): 98–104. doi:10.1016/j.cognition.2009.06.006.

Embodied communication: Speakers' gestures affect listeners' actions

Susan Wagner Cook

The University of Iowa

Michael K. Tanenhaus

The University of Rochester

Abstract

We explored how speakers and listeners use hand gestures as a source of perceptual-motor information during naturalistic communication. After solving the Tower of Hanoi task either with real objects or on a computer, speakers explained the task to listeners. Speakers' hand gestures, but not their speech, reflected properties of the particular objects and the actions that they had previously used to solve the task. Speakers who solved the problem with real objects used more grasping handshapes and produced more curved trajectories during the explanation. 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, and this behavior was related to the particular gestures that were observed. These findings demonstrate that hand gestures are a reliable source of perceptual-motor information during human communication.

Speakers and listeners have access to a wide variety of information about the world, some of which does not readily lend itself to representation in language. A paradigmatic example is the difficulty speakers have in communicating information about how to perform actions in the world -- for example, explaining how to tie one's shoes or ride a bicycle. When people speak, however, they also often gesture with their hands. These gestures may be particularly well-suited for the representation and communication of information about performing actions in the world. However, given that this sort of procedural knowledge is often not linguistically accessible (Willingham, Nissen & Bullemer, 1989), it is also possible that detailed procedural knowledge of a task is not active when people are communicating about a task, especially when they are not overtly performing any actions or making reference to co-present objects.

If speakers do encode action information in their gesture, one further question is whether listeners are sensitive to this information. Outside the context of speech, observers are sensitive to subtle differences in action dynamics when observing overt actions (e.g. Runeson & Frykholm, 1981; Heyes & Foster, 2002). Moreover, a growing body of research suggests that perceptual-motor representations are involved in language processing (Glenberg & Kaschak, 2002; Glover, Rosenbaum, Graham, & Dixon, 2004; Kaschak et al., 2005; Richardson, Spivey,

© 2009 Elsevier B.V. All rights reserved

Address for Correspondence: Susan Wagner Cook Department of Psychology E11 Seashore Hall Iowa City, IA 52242 USA susan-cook@uiowa.edu 319.384.1815.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Barsalou, & McRae, 2003; Zwaan, Stanfield, & Yaxley, 2002). Thus, it is possible that listeners are using action information from gesture to further inform their construction of meaning. However, previous work on listener's sensitivity to hand gestures has focused on listeners' responses to gestures with qualitatively different handshapes, trajectories, and/or locations, leaving open the question of whether listeners are sensitive to fine-grained procedural information in gesture.

Models of gesture production make differing predictions about whether or not action information should be preferentially encoded in gesture. Models in which hand gestures emerge out of interaction between visual-manual and linguistic forms of representation during communication (e.g. McNeill, 1992, Kita & Özyürek, 2003) predict that action information expressed in gesture should also impact the accompanying speech. In contrast, models that propose that information is distributed across speech and gesture (e.g. Beattie & Shovelton, 2006; de Ruiter, 2007) predict that information that readily lends itself to manual representation should be expressed in gesture without impacting the accompanying speech.

At present, models of gesture comprehension have not been fully specified, given that the current debate largely focuses on whether or not gestures serve communicative functions. However, if listeners are sensitive to quantitative differences in action information in the context of speech, this would provide support for the hypothesis that listeners are recruiting information from their perceptual-motor system in service of constructing meaning, and that gesture can be a source of such information. Alternatively, listeners may be sensitive to contrasting features of gestures, without being sensitive to the particular parameters with which these features are produced. On this account, we would not expect listeners to be affected by small differences in gesture form, as long as those gestures shared movement features.

The present research was designed to address three unresolved questions about the interplay between language and gesture in communication. First, are specific action plans activated when speakers describe a problem with a goal structure that is independent of a particular action procedure. Second, if action information is activated, will it be realized in both speech and gesture or only in gesture. Third, are listeners sensitive to any fine-grained procedural information that is potentially available.

To explore these questions, we used the Tower of Hanoi task. Declarative knowledge of the Tower of Hanoi task emerges with experience with the task, and is not tied to any particular procedure for implementing the solution. This knowledge consists of a series of goal states that, if achieved, will lead to the solution. Our hypothesis was that speakers would express this declarative knowledge in their speech, regardless of the particular actions that were used to acquire the knowledge. We further predicted that speakers would express information about the particular actions that they had performed in their gesture, and that listeners would be sensitive to this action information in gesture.

Methods

Fourteen pairs of participants were included in the current study. Data from an additional two pairs in each condition were eliminated due to irregularities in the experimental procedure. The speaker from each pair first solved the Tower of Hanoi problem, while the listener waited in another room. In this problem, a stack of disks, arranged with the largest on the bottom to the smallest on the top, sits on the leftmost of three pegs, and the goal is to move the stack to the rightmost peg, moving only one disk at a time, without ever placing larger disks on top of smaller ones. Speakers solved the problem using either heavy metal disks on wooden pegs or cartoon pictures on a computer screen. The real objects were four weights (0.6, 1.1, 2.3 and 3.4 kilograms) resting on a 22.86 × 76.2 cm board with pegs at 12.7, 38.1 and 63.5 cm. The

computer objects were presented on a 45.7 cm computer monitor with resolution 480 × 640. 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, speakers solved the four-disk problem a second time. They then moved to a different room, where they explained the four-disk problem to the other participant, the listener. After the explanation, listeners returned to the original room and solved and explained the four-disk problem twice. All listeners solved the task on the computer (speakers in both conditions 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 moved the apparatus out of view while the two participants were in the explanation room. Speakers' and listeners' mouse trajectories were tracked while they solved the problem on the computer (Spivey, Grosjean, & Knoblich, 2005).

Coding

Speech was transcribed for subsequent analysis. Gestures were coded for the hand shape and trajectory to explore possible differences predicted from the affordances of the objects used in the two conditions.¹ All gestures depicting movement of the disks from one peg to another were identified in the video, and each gesture was coded for the handshape of the moving hand(s) (Grasping/Handling or NonGrasping/NonHandling).² Inter-rater reliability of this judgment across two independent observers was .85. Gesture trajectory was coded from the video record of the explanation. To calculate the trajectory, the video of each gesture was transformed into a series of still images sampled at a rate of 10 per second. The position of each hand was then coded from the still images by recording the location of the knuckle of the index finger of each hand in each image using screen coordinates. For marking the hand position, the mean distance between the points marked by the two independent observers was 13 pixels with a range of 0 to 48 pixels (screen resolution was 990×1440). Judgments of the more experienced coder were used in all analyses.³

Results

Spoken explanations

To explore how speakers communicated this information to listeners, we investigated both speech and gesture produced during the explanation. We expected that action information might be particularly likely to be encoded in the verb, and so we first analyzed the verb used across conditions. We first identified the stem for each verb and then analyzed across stems. In multilevel models with condition and stem as fixed factors, words spoken (log-transformed) as a covariate, and participant as a random factor, there were no effects of condition on either the likelihood of producing particular verbs, or the frequency with which particular verbs were spoken. Speakers in both conditions overwhelmingly tended to use a single verb, “move”, to describe the action.

However, speakers may have expressed differences between conditions more generally in speech. Accordingly, we further compared encoding in speech using the overall distribution of word types and the frequency of tokens produced by speakers across conditions. There were 286 different word types produced by speakers. Speakers in the real object condition produced

¹The complex dynamics of gestures and the relatively limited methods of measuring these dynamics employed here make it likely that additional features of participants' gestures also reflect differences across conditions.

²One participant in the computer condition was eliminated from the gesture coding, because the gestures were produced primarily perpendicular to the camera and so were unable to be coded from the video.

³The pattern of findings does not change if the judgments of the less experienced coder are used in the analyses.

an average of 66 types and speakers in the computer condition produced an average of 82 types ($t(12)=1.17$, ns). The likelihood of and frequency of producing each content word in each condition was examined using multilevel logistic regression models with condition and stem as fixed factors, words spoken (log-transformed) as a covariate, and participant as a random factor. There were no significant interactions between word and condition, indicating that there were not any particular word stems that speakers in one condition were more likely to produce, or produced more frequently. Finally, an additional group of 18 participants given only the audio were unable to guess which condition a given speaker had participated in (54% correct, no different from chance (50%), $t(17)=1.2$, ns). Thus, it appeared that, despite their differing action experience, speakers in the real objects and the computer condition used similar spoken language to describe how to solve the Tower of Hanoi.

Gesture during the explanation

We next explored speakers' gestures. All speakers ($N=14$) depicted the movement of the disks using hand gestures in conjunction with their speech. There were no differences between conditions in the rate of gestures per word produced during the explanation (Computer: .051 vs. Real: .051).

We investigated encoding of information in gesture by comparing hand shapes and trajectories of motion across conditions. To examine difference in hand shape, we used multilevel logistic regression with the likelihood of a gesture having a grasping hand shape (versus a non-grasping handshape) as the dependent measure, condition as a fixed factor and subject as a random factor. Speakers in the real-objects condition used a larger proportion of grasping hand shapes, in comparison with speakers in the computer condition (Real objects: 80% grasping gestures versus Computer: 49%, Wald $z=1.91$, $p=.05$).

To investigate differences in trajectory, we first extracted the X and Y position of the moving hand for each gesture over time, and then transformed the X coordinates of the coded hand location so that the coordinates ranged from 0 to 1, and the Y coordinates so that the minimum position was 0, for each gesture.⁴ Because the gestures produced by speakers were primarily curved movements depicting the trajectory of the disks, we modeled the position of each hand using a maximum likelihood quadratic model to predict the y-coordinate, given condition, the x-coordinate of the hand, the quadratic of the x-coordinate, and the condition under which the speaker had solved the problem. We used a Markov Chain Monte Carlo sample to estimate the p values of coefficients. There was a reliable quadratic component associated with hand position in each condition (Computer: $\beta=-316.8$, $p<.0001$; Real: $\beta=-438.8$, $p<.0001$), indicating that participants in both conditions produced gestures with curved trajectories. However, as can be seen in Figure 2, the size of the quadratic component was significantly greater in the real-objects condition, leading to gestures with more curved trajectories. Consistent with this observation, the model including an interaction with condition provided a better fit to the data than a model without this interaction term ($\chi^2(1)= 8.55$, $p=.003$). Moreover, differences in gesture were perceptible to observers. An additional group of 18 participants given only the video were able to reliably guess which condition a speaker had participated in at better than chance performance (58% correct, $t(17)=2.79$, $p=.012$, two-tailed). Thus, unlike the speech, the gestures produced by speakers reflected differences in the materials used to solve the problem - speakers who solved the problem with real objects used more grasping hand shapes and gestured with higher, more curved trajectories.

⁴For gestures using both hands, we used data from the right hand. Because both hands moved in synchrony in the gestures produced with two hands in this experiment, the transformed position of the each of the hands over time is the same.

Listeners' movements

We examined the mouse trajectories produced by listeners when they subsequently solved the problem to determine whether listeners extracted differential information across conditions. In order to compare trajectories 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 (the minimum needed to solve the problem) produced by each participant were included in the analysis. We used a maximum likelihood quadratic model to predict the height of the mouse, given the x-coordinate of the movement, the quadratic of the x-coordinate, and the condition under which the speaker had solved the problem. The dependent measure was the y-coordinate and participant was a random factor. We used a Markov Chain Monte Carlo sample to estimate the p values of the interaction coefficients. The function fitting the mouse trajectories had a significantly larger quadratic 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 (Computer: $\beta = -174.7$, $p < .0001$; Real: $\beta = -261.1$, $p < .0001$; See Figure 3).⁵ ⁶ Moreover, the model including an interaction with condition provided a better fit to the data than a model without this interaction term ($\chi^2(1) = 182.0$, $p < .0001$). Thus, listeners' mouse trajectories differentially reflected the procedural knowledge experience of the speaker.

Finally, we explored whether cues from speakers' gesture were related to listeners' behavior. For each speaker, we estimated an average trajectory of motion when gesturing using a quadratic regression model and then calculated the average gesture height for each speaker using the value predicted halfway through the gesture. We also calculated the number of gestures produced with a grasping hand shape for each speaker. For each listener, we estimated the average trajectory of mouse motion using a quadratic model, and similarly calculated the listeners' mouse height using the value predicted halfway through the average movement. We then investigated whether parameters of speakers' gestures were related to the height of listeners' mouse movements using a regression model with speakers' average gesture height and number of grasping gestures as predictors, and listeners' mouse height as the dependent measure. As depicted in Figure 2, there was a significant positive relation between the average height of a particular speakers' gesture and the height of the listeners' subsequent mouse movements ($\beta = .7$, $t(9) = 2.813$, $p = .02$). The effect of number of grasping gestures did not reach significance ($\beta = 4.7$, $t(9) = 1.2$, ns).

Discussion

Our results demonstrate that embodied action-plans are activated when speakers describe the goal structure of a task. Moreover, speakers' hand gestures, but not their speech, reflected their procedural knowledge of the task, demonstrating that gesture can be a vehicle for the expression of information that is unlikely to be expressed in the accompanying speech. This clear division of labor supports models in which information can be expressed in gesture without impacting the accompanying speech (e.g. Beattie & Shovelton, 2006; de Ruiter, 2007).

Furthermore, procedural information from gesture was incorporated into listeners' subsequent behavior. These findings go beyond previous demonstrations that listeners glean information from gesture to reveal that listeners can extract fine-grained perceptual-motor information from gesture during naturalistic communication.

⁵Additional reliable fixed effects in the model were as follows: There was a reliable linear component to the movements ($\beta = 163.0$, $p < .0001$), and the linear component interacted with condition (Real: $\beta = 91.1$, $p < .0001$).

⁶Listeners were equally facile at solving the task across conditions. There was not a statistically significant 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, $t(12) = .83$).

These findings have important implications for theories of gesture production as well as gesture comprehension. In addition to online sensitivity to syntactic packaging in the accompanying speech (Kita & Özyürek, 2003; Kita et al, 2007), gestures have access to detailed information from the perceptual-motor system that does not influence the accompanying speech. Speakers in the current study who had solved the problem with real objects reproduced features of their motor movements in their gesture. This suggests that, during communication, speakers were not only activating the particular goal of their movement, which were expressed in the concurrent speech (e.g. “Move the largest disk all the way over”), but also the specific motor plan that would accomplish this goal, which was expressed in the concurrent gesture (e.g. grasping and lifting or clicking and dragging). Although it had been hypothesized that gestures emerge from perceptual and motor simulations during communication (Hostetter & Alibali, 2008; Kita, 2000; McNeill, 1992; Streeck, 1996), it was not known whether speakers provide reliable perceptual-motor information in their gesture when physical objects are not available in the immediate environment. Speakers could also derive their gestured representations from more abstract representations of task goals.

Moreover, the relation between the trajectory of the observed gesture and the trajectory of the subsequent mouse movement suggests that imitative learning of procedural knowledge can occur even when actions are represented, rather than performed overtly. When processing sign language, signers often ignore iconic features of signs - the neural activation associated with processing signs, even pantomime-like signs, is similar to that associated with processing non-iconic spoken words and different from that observed when processing iconic pantomimes (Emmory et al., 2004). Yet, outside the context of speech, observers can use quantitative features of another person's motor behavior to make inferences about the object that person is interacting with. For example, observers can detect the weight of a box that is being lifted by another person (Runeson & Frykholm, 1981). Here we demonstrated that even in the context of spoken communication, listeners are sensitive to small changes in movement dynamics, suggesting that, like speakers, listeners may be constructing perceptual-motor simulations during communication.

There is considerable evidence that listeners are affected by speakers' gesture (e.g. Cassell, McNeill, & McCullough, 1999; Goldin-Meadow, Kim, & Singer, 1999; Graham & Argyle, 1975; Holle & Gunter, 2007; Kelly, Kravitz, & Hopkins, 2004; Özyürek, Willems, Kita & Hagoort, 2007; Silverman et al., in press; Wu & Coulson, 2005). However, much of this work has presented gesture under carefully controlled conditions, using targeted assessments of information uptake. Here speakers' gesture affected listeners' behavior during spontaneous communication (see Goldin-Meadow and Singer, 2003, for related findings). Moreover, these findings provide the first evidence that listeners are sensitive to quantitative differences in gesture form. Listeners in all conditions saw grasping and pointing hands moving in curved trajectories, what varied across conditions was the probability of grasping handshapes, and the amount of curvature in the trajectories. These features did not go unheeded, but rather were related to the subsequent behavior of the listeners. Thus, in addition to being sensitive to qualitative differences in gesture shape, trajectory or location, listeners can use quantitative information from gesture to extract information about the world.

These findings are consistent with the hypothesis that, when interpreting language, even abstract language, listeners activate perceptual and motor information (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). As this example demonstrates, the activation of perceptual and motor representations in listeners has been inferred primarily from facilitation and/or decrements in performance on secondary perceptual-motor tasks like picture

verification or lever pulling. These secondary tasks may be reorganizing processing by activating the very systems under investigation. For example, instructions to move one's own body can facilitate recognition of bodily postures in others, by activating aspects of one's body schema that would not be recruited without instructions (Reed & Farah, 1995). Here we provide evidence that perceptual-motor representations are activated during human communication using data from naturalistic behavior of speakers and listeners, revealing that these effects are not artifacts of the particular secondary tasks used to measure perceptual and motor processing.

In sum, our hands are clearly a valuable resource for communicating information about performing actions in the world that is not expressed in concomitant speech; human communication is not only do as I say, but also do as I do.

Acknowledgements

We thank N. Cook, M. Hare, K. Housel, S. Packard., A.P. Salverda, D. Subik and T. Truelove for assistance in various aspects of executing the experiment and R. Achtman, R. Aslin, N. Cook, K. Housel and R. Webb for helpful comments on the manuscript. This work was supported by NIH grant HD-27206.

References

- Beattie G, Shovelton H. When size really matters: How a single semantic feature is represented in the speech and gesture modalities. *Gesture* 2006;6(1):63–84.
- Cassell J, McNeill D, McCullough K-E. Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics & Cognition* 1999;7(1):1–34.
- de Ruiter JP. Postcards from the mind: The relationship between speech, imagistic gesture, and thought. *Gesture* 2007;7(1):21–38.
- Emmorey K, Grabowski T, McCullough S, Damasio H, Ponto L, Hichwa R, Bellugi U. Motor-iconicity of sign language does not alter the neural systems underlying tool and action naming. *Brain and Language* 2004;89(1):27–37. [PubMed: 15010234]
- Glenberg AM, Kaschak MP. Grounding language in action. *Psychonomic Bulletin & Review* 2002;9(3): 558–565. [PubMed: 12412897]
- Glover S, Rosenbaum DA, Graham J, Dixon P. Grasping the meaning of words. *Experimental Brain Research* 2004;154(1):103–108.
- Goldin-Meadow S, Kim S, Singer M. What the teacher's hands tell the student's mind about math. *Journal of Educational Psychology* 1999;91(4):720–730.
- Goldin-Meadow S, Singer MA. From children's hands to adults' ears: Gesture's role in the learning process. *Developmental Psychology* 2003;39(3):509–520. [PubMed: 12760519]
- Graham JA, Argyle M. A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology* 1975;10(1):57–67.
- Heyes CM, Foster CL. Motor learning by observation: Evidence from a serial reaction time task. *The Quarterly Journal of Experimental Psychology: Section A* 2002;55:593–607.
- Holle H, Gunter TC. The Role of Iconic Gestures in Speech Disambiguation: ERP Evidence. *Journal of Cognitive Neuroscience* 2007;19(7):1175–1192. [PubMed: 17583993]
- Hostetter AB, Alibali MW. Visible embodiment: Gestures as Simulated Action. *Psychonomic Bulletin and Review* 2008;15(3):495–514. [PubMed: 18567247]
- Kaschak MP, Madden CJ, Theriault DJ, Yaxley RH, Aveyard M, Blanchard AA, et al. Perception of motion affects language processing. *Cognition* 2005;94(3):B79–B89. [PubMed: 15617669]
- Kelly SD, Kravitz C, Hopkins M. Neural correlates of bimodal speech and gesture comprehension. *Brain and Language* 2004;89(1):253–260. [PubMed: 15010257]
- Kita S, Özyürek A, Allen S, Brown A, Furman R, Ishizuka T. Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes* 2007;22(3):1–25.
- Kita, S. How representational gestures help speaking. In: McNeill, D., editor. *Language and gesture*. Cambridge University Press; Cambridge: 2000. p. 162-185.

- Kita S, Özyürek A. 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* 2003;48:16–32.
- McNeill, D. *Hand and mind: What gestures reveal about thought*. The University of Chicago Press; Chicago: 1992.
- Özyürek A, Willems RM, Kita S, Hagoort P. On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience* 2007;19(4):605–616. [PubMed: 17381252]
- Reed CL, Farah MJ. The Psychological Reality of the Body Schema - a Test with Normal Participants. *Journal of Experimental Psychology-Human Perception and Performance* 1995;21(2):334–343. [PubMed: 7714475]
- Richardson DC, Spivey MJ, Barsalou LW, McRae K. Spatial representations activated during real-time comprehension of verbs. *Cognitive Science* 2003;27(5):767–780.
- Runeson S, Frykholm G. Visual perception of lifted weight. *Journal of Experimental Psychology - Human Perception and Performance* 1981;7(4):733–740. [PubMed: 6457088]
- Silverman L, Campana EM, Bennetto L, Tanenhaus MK. Speech and gesture integration in high functioning Autism. *Cognition*. (in press)
- Spivey MJ, Grosjean M, Knoblich G. Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences* 2005;102(29):10393–10398.
- Streeck J. How to do things with things. *Human Studies* 1996;19(4):365–384.
- Willingham DB, Nissen MJ, Bullemer P. On the development of procedural knowledge. *Journal of experimental psychology: Learning, memory, and cognition* 1989;15:1047–60.
- Wu YC, Coulson S. Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology* 2005;42:654–667. [PubMed: 16364061]
- Zwaan RA, Stanfield RA, Yaxley RH. Language comprehenders mentally represent the shape of objects. *Psychological Science* 2002;13(2):168–171. [PubMed: 11934002]

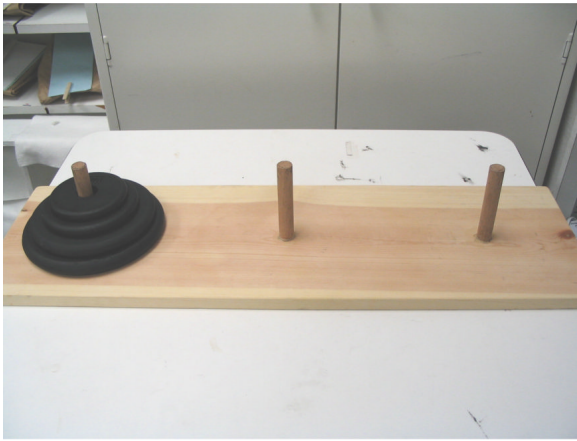
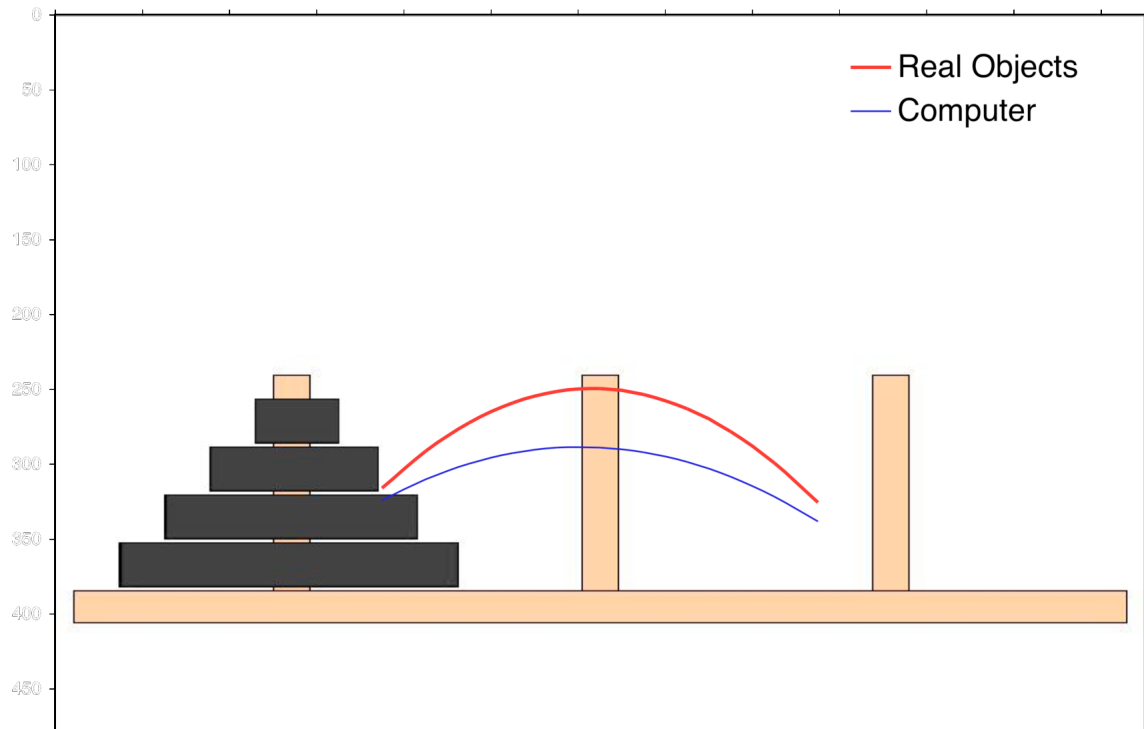
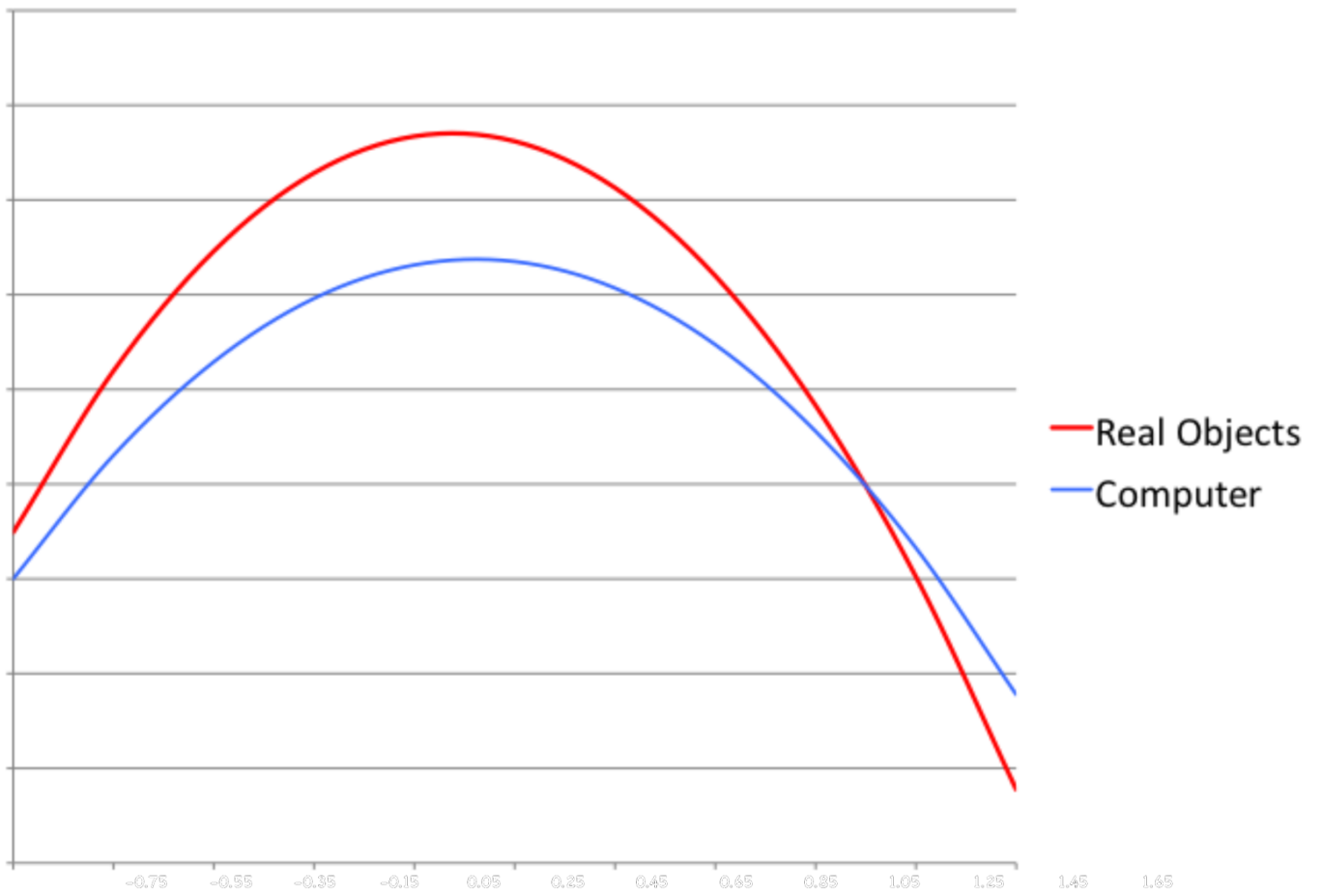


Figure 1.
The Towers of Hanoi used in the Real Objects and Computer conditions.



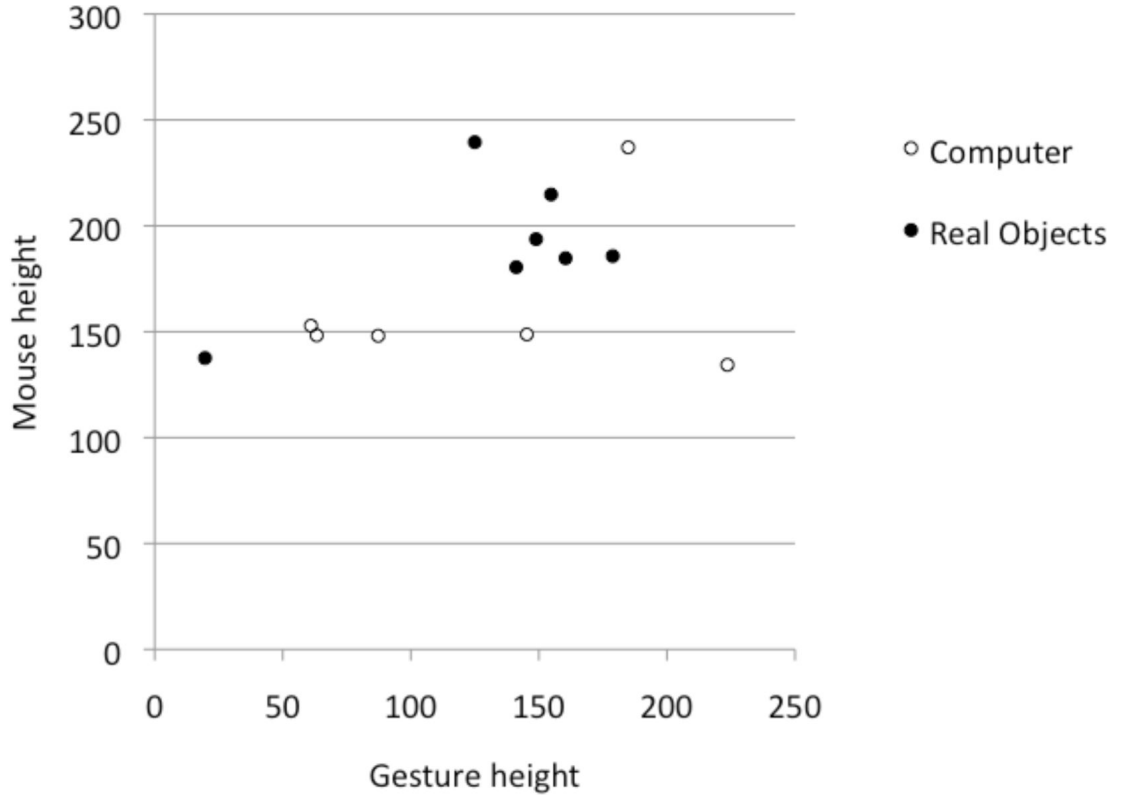


Figure 2. Differences in listeners' movement of the computer objects as a function of the motor experience of the speaker each listener observed. (a) Gesture trajectories predicted from the position of speakers' hands over time. The x and y axes are video image coordinates. (b) Mouse trajectories predicted using data from the first 15 moves made by the listener, superimposed on the computer display. The x and y axes are screen coordinates. (c) The relation between speakers' gesture height and listeners' mouse height. The x and y axes are pixels.