Spring 2012

Can participants extract subtle information from gesturelike visual stimuli that are coordinated with speech without using any other cues?

Marwa Abdalla
University of Iowa

CAN PARTICIPANTS EXTRACT SUBTLE INFORMATION FROM GESTURE-LIKE VISUAL STIMULI THAT ARE COORDINATED WITH SPEECH WITHOUT USING ANY OTHER CUES?

by

Marwa Abdalla

A thesis submitted in partial fulfillment
of the requirements for the Master of
Arts degree in Psychology
in the Graduate College of
The University of Iowa

May 2012

Thesis Supervisor: Assistant Professor Susan Wagner Cook

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

____

MASTER'S THESIS

____

This is to certify that the Master's thesis of

Marwa Abdalla

has been approved by the Examining Committee for the thesis requirement for the Master of Arts degree in Psychology at the May 2012 graduation.

Thesis Committee: ___________________________________
Susan Wagner Cook, Thesis Supervisor

____
Bob McMurray

____
Julie Gros-Louis

____
Patricia Zebrowski

To My Family, for their care and loving support

ABSTRACT

Embodied cognition is the reflection of an organism's interaction with its environment on its cognitive processes. We explored the question whether participants are able to pick up on subtle cues from gestures using the Tower of Hanoi task. Previous research has shown that listeners are sensitive to the height of the gestures that they observe, and reflect this knowledge in their mouse movements (Cook & Tanenhaus, 2009). Participants in our study watched a modified video of someone explaining the Tower of Hanoi puzzle solution, so that participants only saw a black background with two moving dots representing the hand positions from the original explanation in space and time. We parametrically manipulated the location of the dots to examine whether listeners were sensitive to this subtle variation. We selected the transfer gestures from the original explanation, and tracked the hand positions with dots at varying heights relative to the original gesture height. The experimental gesture heights reflected 0%, 25%, 50%, 75% and 100% of this original height. We predicted, based on previous research (Cook in prep), that participants will be able to extract the difference in gesture height and reflect this in their mouse movements when solving the problem. Using linear model for our analysis, we found that the starting trajectory confirmed our hypothesis. However, when looking at the averaged first 15 moves (the minimum to solve the puzzle) across the five conditions, the ordered effect of the gesture heights was lost, although there were apparent differences between the gesture heights. This is an important finding because it shows that participants are able to glean subtle height information from gestures. Listeners truly interpret iconic gestures iconically.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

INTRODUCTION

When people have a conversation, they generally use gestures. Gesture provides a visual representation that is complementary to speech. Accordingly, people may use gesture to communicate information and ideas, particularly when speech cannot effectively communicate that information. In this thesis, I explore the nature of the information communicated via visual information accompanying speech.

Gestures have the potential to encode continuous information about the world. Hand movements can directly express visual and spatial properties of the world. Moreover, the shape of gestures changes over an interaction (Holler & Stevens, 2007; Holler & Wilkin, 2009), demonstrating that gestures can be adjusted according to the communicative context. MacGregor (2003) illustrates the potential for representing continuous information during communication using a scene from a movie, where the actor Jack Lemmon changes a simple toilet cardboard roll into a firecracker, then using his present state to transition into a future state of how he uses the firecracker to blow up the bed bunk of his captain. A quote from his paper clarifies the potential for iconic representation in American Sign Language compared with speech:

> The word 'Bam', then, seems to encode a relatively small explosion, as opposed to e.g. 'Boom', which encodes a larger explosion...In this respect, the difference in vowels is iconic: the high back rounded vowel in 'Boom' sounds deeper than the low front vowel in 'Bam', and so seems to represent a larger explosion…These are gradient phenomena: the louder the voice and the more violent the motion, the larger the explosion depicted…There is, however, an important difference between depicting verbs in ASL and onomatopoeic words in a spoken language: since ASL uses a visual modality, what is depicted in ASL tends to be the visual element of the action, while onomatopoeic words in spoken languages depict the auditory element of the action. (p. 82)

In the same way that sound can be used to iconically and gradiently represent the auditory world, gesture can iconically and gradiently represent the visual world.

Speech and gesture express different things (Alibali et al., 2009). Gestures are best for describing things visually; while it would be very cumbersome for a speaker to explain the shape of the United States or the curved trajectory of a leaf falling to the ground in speech, these concepts can be efficiently represented in gesture. If gestures represent the world iconically, gestures should encode continuous information about the visual world. For example, a speaker may represent a ball with two horizontally placed hands that are facing each other's palms, with fingers curved and spread out. The distance between the hands might relate to the actual size of the ball that is being discussed (Beattie & Shevolton, 2006). Yet variation in the distance between the hands is likely to also vary according to a variety of factors, including common ground, the level of arousal of the speaker, and the size of their gesture space, which additionally varies by culture. Given the myriad factors that might influence the shape of a particular gesture, gesture might not be particularly likely to truly iconically represent the world, and thus listeners may not be sensitive to this dimension. How do people interpret information in gesture?

One possibility is that participants are sensitive to continuous information in gesture, and interpret at least some of the information in gesture iconically. The alternative is that information in gesture is interpreted categorically (McNeill, 1992). Emmorey et al. (2003) described categorical perception as "a set of stimuli ranging along a physical continuum are identified as belonging to distinct, bounded categories, and subjects are better able to discriminate between pairs of stimuli that straddle this boundary than pairs that fall within one category or the other." The earliest research on categorical perception was in the domain of speech perception. For example, Liberman et al. (1957) tested participants’ ability to discriminate between speech sounds using an ABX task. In this task, A and B are different stimuli, while X is the same as A or B. Participants’ job is to figure out whether X is the same as A or B. They found that participants are better at discriminating sounds at the phoneme boundaries (between b and d, or d and g), than within phoneme categories (within variations of b, d, or g) (1957), suggesting that phonemes are perceived categorically. Categorical perception is not unique to speech or auditory perception but can also be seen in visual processing of facial expressions. McCullough & Emmorey (2009) manipulated facial expressions on a continuum. Hearing non-signers were able to discriminate facial expressions across categorical boundaries (not within categories) better the Deaf signers. However, Deaf signers were able to discriminate between linguistic facial expressions, since they are used more in ASL than in a social context.

There is a wide literature demonstrating that listeners are sensitive to information in gesture; however, this literature has not directly explored the question of whether or not visual information in gesture is interpreted iconically. Participants might attend to the features of the gesture without directly mapping information in gesture to the world. That is to say, a participant viewing a ball gesture might be expected to treat varying sizes of ball gestures as representing either large or small balls, without using the fine-grained information in gesture as a source of information about the actual size of the gesture. When the two hands are close together, participants could infer a small ball, and when the two hands are far apart, participants infer a large ball. Similarly, participants might treat some gestures as representing curved trajectories and other trajectories as representing straight trajectories without attending to the specific amount of curvature expressed in gesture.

Participants are clearly influenced by information in gesture. Thompson & Massaro (1986) conducted an experiment where they manipulated ba and da on a speech continuum combined with gesture pointing to either ball or doll. The participants were in three groups, speech only, gesture only, and speech and gesture. Using speech or gesture, participants had to figure out what was the referred object. It turns out that many of the participants were influenced by gestures in inferring which object was referred to, particularly when speech was ambiguous.

In another example, Goldin-Meadow & Sandhofer (1999) tested adults’ ability to read children’s gestures. They showed that participants were able to glean information that were unique to gestures by comparing participants’ responses to vignettes that had the same speech, but one that had gestures and one that did not or had gestures that were incongruent with speech. They found that gestures incongruent with speech hindered adults' comprehension of the children's explanations, but gestures congruent with speech merely did not enhance comprehension above no gesture. However, they never tested if adults were able to distinguish between the children's different gesture types.

In a separate study, Cassell et al. (1998) explored how mismatches in the information in speech and gesture affected the listener's retelling of a story narrated by a speaker. One of these mismatches was the speaker's perspective of the iconic gesture (self or character in the story). For example, the speaker is talking from one character's point of view, yet the gesture implies that the speaker is talking from another character's view. The researchers found that listeners incorporated information from these gestures mismatches in their story retellings. However, this study did not test if participants were able to distinguish between gesture types, such as iconic and metaphoric gestures, let alone delve into the idea that participants might have the ability to distinguish variation within one gesture, such as height or size information in an iconic gesture.

Information from gesture also helps listeners resolve ambiguities in speech (Kelly et al., 1999; Obermeier et al., 2011). In one example, Kelly et al. (1999) conducted a study to understand people's understanding of indirect requests, both with and without gesture. They set up two conditions where participants watched a video of professional actors acting out a script. The script ended with an ambiguous indirect request statement. The speech and gesture condition included a deictic gesture accompanying the statement, and the speech condition did not. They found that participants understood the indirect requests better with gesture than without.

Other research studies manipulated the relationship between gesture and speech, and found when a gesture is semantically or temporally more aligned with the speech, participants are better able to integrate gesture and speech to get the correct information (Habets et al., 2011; Treffner et al., 2008). For example, in an ERP study, Habet et al.'s

(2011) manipulated gesture-speech synchrony by delaying the onset of speech to test participants' ability to integrate speech and gesture. They found that when there was temporal overlap between speech and gesture, even though speech occurred simultaneously or delayed after gesture, participants showed high N400 amplitudes for mismatching gesture. However, the effect of the speech delay has its limits. When speech did not have any temporal overlap with gesture, participants did not display the high N400 amplitudes.

There is also some evidence that participants cannot avoid being influenced by visual gestures even when they are not helpful. In one study, the direction of the eye gaze was pitted against gestures (i.e. looking up and pointing down) in conditions with congruent and incongruent speech (Langton & Bruce, 2000). Participants had to quickly respond to the speech while seeing the congruent/incongruent visual eye gaze and gesture information. They found that participants were faster to respond appropriately when the gestures were congruent with the voice, even when eye gaze was incongruent with gesture. When gestures were incongruent with eye gaze, participants took longer, even when eye gaze was congruent.

If listeners are sensitive to gesture, they may also be sensitive to iconic and continuous information in gesture. Previous research showed that size information is encoded iconically in gestures; however, it is not known whether listeners are sensitive to this information during communication. Beattie & Shovelton (2006) conducted a study to focus on gesture size in retelling stories. Participants watched three cartoon stories, and had to retell them to their listeners in a random order on a wall in front of them. The listeners were judges and scored parts of these retellings. After coding for instances of size information in speech, gesture, and speech-gesture, they found that one of the most salient iconic gestures used in the retellings was the size information of the ball. It was only found in gesture, never found in speech or in instances where both speech and gesture mentioned the size of the ball. Any extremely crucial size information was demonstrated in gesture. For example, in one of the cartoons used in this experiment, it was a story of a dog trapping kittens inside a ball and playing with it. The dog used this ball quite frequently in the story; thus, it was important for the listeners to know the size information of the ball. The researchers found that participants rated size information that was only found in gestures as most important to the story.

Outside the context of speech, researchers have found that participants are able to glean subtle information from motor cues that are similar to gesture (Runeson & Frykholm, 1981). Runeson & Frykholm (1981) conducted a study where participants had to judge the weight of the box from just viewing a person picking it up without any auditory cues. They conducted two experiments. In the first one, they lined the actor and the box with retroreflective tape. The actor was recorded picking up the box in the dark. There were five different weights: 4, 10, 16, 22, and 28 kg. The actor did not know the weight of the box ahead of time. Participants were asked to watch the actor picking up different boxes and judge the weights of the boxes. They were able to accurately judge the weight of the boxes given the visual lifting information. In the second experiment, participants were live observers while actors carried different weights. They changed the weight to 4, 9, 14, 19, and 24 kg to minimize risk of injury. They found that participants observing someone carrying weights were almost as accurate at judging the weight as carrying the weight boxes by themselves.

Evidence from sign language also suggests that participants may be sensitive to to *continuous* information in visual gestures. One previous research study found that participants sometimes view ASL signs categorically and sometimes on a continuum (Emmorey et al., 2003). In this study, hearing non-signers and Deaf signers were tested on their perception of differing ASL signs. The signs were manipulated so that important features varied continuously. In this study, the signers showed categorical perception of the hand information, but only for some features. For other features, both signers and non-signers interpreted the information in a gradient manner. This means that visual information presented via hands can be interpreted in both ways. Importantly, experience seems to drive categorical perception – only viewers with experience interpreted stimuli categorically. Because hearing listeners have extensive experience with gesture, they might also interpret gestures categorically.

Other previous research has indirectly suggested sensitivity to continuous information in gesture (Cook & Tanenhaus, 2009). This study demonstrated that listeners glean information from gesture when speakers explained how to solve the Tower of Hanoi puzzle. Speakers either solved the puzzle in its physical form- heavy metal disks on wooden pegs, or as a cartoon game on the computer. Participants explained how to solve the problem to the listeners, and listeners had to go and solve the puzzle on the computer. When the speakers from both groups explained to the listeners how to solve the puzzle, they did not provide any verbal information about the manner in which they lifted the disks to place them in different pegs. However, speakers did provide information about manner in their gestures; speakers who solved the problem with real objects produced gestures with more curved trajectories. Moreover, listeners showed more curved mouse trajectories when they listened to the speakers who solved the physical form of the puzzle than those who solved the puzzle on the computer. In addition, listeners' average trajectory of the mouse movement was positively correlated with the trajectory of the speaker's gestures that the listener had observed. This shows that listeners picked up on curved information from the gestures.

Although there is evidence that people are influenced by gesture and that gestures have the potential to convey gradient information, prior research on gesture has not directly explored the potential for listener sensitivity to continuous, gradient information in gesture, nor controlled for tasks that would elicit gradient information. After Liberman and others showed that participants perceive speech categorically, another group of researchers argued that research paradigms that were used pushed for the categorical effect in perception, as opposed to truly showing perception in a continuous manner

(Pisoni & Lazarus, 1974; Hary & Massaro, 1982; Pastore et al, 1977). For example, Pisoni & Lazarus (1974) demonstrated that another task is a better measure of continuity perception in speech than the task Liberman used. Participants compared a pair of stimuli that are the same with a pair of stimuli that are different, and had to figure out which of these two pairs were different. Researchers showed that participants did not show a strong change at the categorical boundary, and inferred that having a task where phonemes are compared rather than committed to memory helps participants rely on relative instead of absolute auditory differences in phonemes. Moreover, Hary & Massaro (1982) simplified their experimental task even further by having participants compare two stimuli and state if they were the same or different. They still found the same results. In another account, Pastore et al. (1977) conducted an experiment where participants had to compare two different tones in a silent background in comparison to while listening to a reference tone in the background (as cited in Hary & Massaro, 1982). They found that when there was a reference tone, participants displayed a categorical effect in their discrimination of the test tones. However, merely comparing the two test tones alone displayed a continuous effect. Thus, these studies demonstrated that the task that is given to the participants influences whether they appear to perceive experimental stimuli categorically or in a gradient fashion.

We have the same problem in gesture research. The tasks used to explore listener sensitivity to gesture are not sensitive to potential iconicity in gesture. First, tasks used usually require categorical judgments, either from the coders (Galati & Brennan, 2010) or from the participants (Treffner, 2008). Using categorical measures makes it less likely that researchers will detect gradient effects of gesture on behavior. To test for sensitivity to continuous information, we need to look be sure that our measure of listener information can vary continuously. Dale et al. (2007) showed gradiency in motor responses using mouse movements. Researchers pitted descriptive words to see the effects of the semantic information on participants' mouse movements. They found that participants' mouse movements reflect a curved path to the correct choice between competing words, because they would be hesitant to make a choice at first. Thus, they would stay in the middle of two choices, until the ambiguity disappeared. For example, they tested whether participants were more likely to describe the whale as a mammal or as a fish, and how their mouse movements reflected their categorical or continuous perception of the animal. Interestingly, participants also showed hesitancy between the choices bird and mammal for whale, which means excluding whale, participants' movements had an attractor state to bird because birds are mammals as well. An earlier study by Spivey et al. (2005) tested phonological competitors between a target and a cohort word (i.e. candle and candy), as speech unfolded over time. They found that once phonological ambiguity between the two words disappeared (the ambiguity being the first part of the word "cand"), the slow and hesitant mouse trajectory quickly marked the correct choice.

The present study endeavors to use mouse tracking to see if participants glean continuous information from visual information in Tower of Hanoi. The Tower of Hanoi is a good task that elicits a lot of gestures when participants explain the solution to the problem to their listeners (Garbar & Goldin-Meadow, 2002; Cook & Tanenhaus, 2009; Beilock & Goldin-Meadow, 2010). The Tower of Hanoi is a puzzle that has a wooden board and three wooden pegs spaced evenly from each other in a straight line. One of those pegs has a tower of metal disks that arranged from largest at the bottom to smallest at the top. In our case, the goal of this puzzle is to move the tower of disks from the left peg to right peg abiding by two rules. The first rule is that only one disk can be moved at a time, and that is the topmost disk on any peg can be moved to another peg. The last rule is that a larger disk cannot be placed on top of a smaller one, only smaller disks can be places on larger ones.

A second problem with prior work is that the nature of the information included in gestures accompanying speech has not been varied continuously. For example, looking at

Treffner et al.’s (2008), we find that they use the same deictic gesture, while varying the timing of the gesture with speech. They did not vary the features of the deictic gesture. Similarly, in Beattie & Shevolton (2006) study, they did not vary the size of the salient object in the story (like different sizes of a ball or balls) or directly measure listener processing of the relevant size information.

In order to assess the potential for listeners to be affected by continuous information in gesture, we need to parametrically manipulate the information presented to listeners. We decided to use point light representations of gesture in order to ensure that the relevant features of the gesture varied appropriately. Previous research (Cook in prep), using point light displays to represent gesture in blank background video found that participants were able to detect the difference between high and low gestures. They reflected this information in their mouse trajectories because they were able to pick up on motion of the balls and extract the gesture height information. The present study takes this idea further, in that we seek to understand whether participants will be able to pick up even more subtle differences between gesture heights on a continuum or not, and will they reflect that information appropriately in an orderly fashion through their mouse trajectories.

This study goes beyond previous studies in that we varied the gesture height parametrically. The gestures that we manipulated are the transfer gestures, which are gestures where a person is moving an object from one point to another point in space. We controlled for the amount of speech produced by the speaker in all of our conditions. We also controlled for the amount of gestures produced by the speaker in all of our conditions. In addition, we methodically controlled for the variation of gesture heights. Cook & Tanenhaus (2009) did not control for the words used and amount of words in their experiment. They also did not control for the amount of gestures or the types of gestures produced by the speakers, nor did they control for varying gesture heights when speakers explained their solutions to the listeners. Therefore, it is hard to parse out what could be the probable cause for their results. We produced five different gesture heights to simulate gesture heights on a continuum, and we predict that participants will be influenced by the variation. Moreover, we predict that they will subsequently show this influence in their behavior.

METHODS

Participants:

150 University of Iowa undergraduate students participated in this study. Thirteen were discarded due to experimenter error, technical problems, or participants were not able to solve the puzzle. After discarding participants, there were 89 females and 48 males that participated in the study. Participants were given research credit for their participation.

Stimuli:

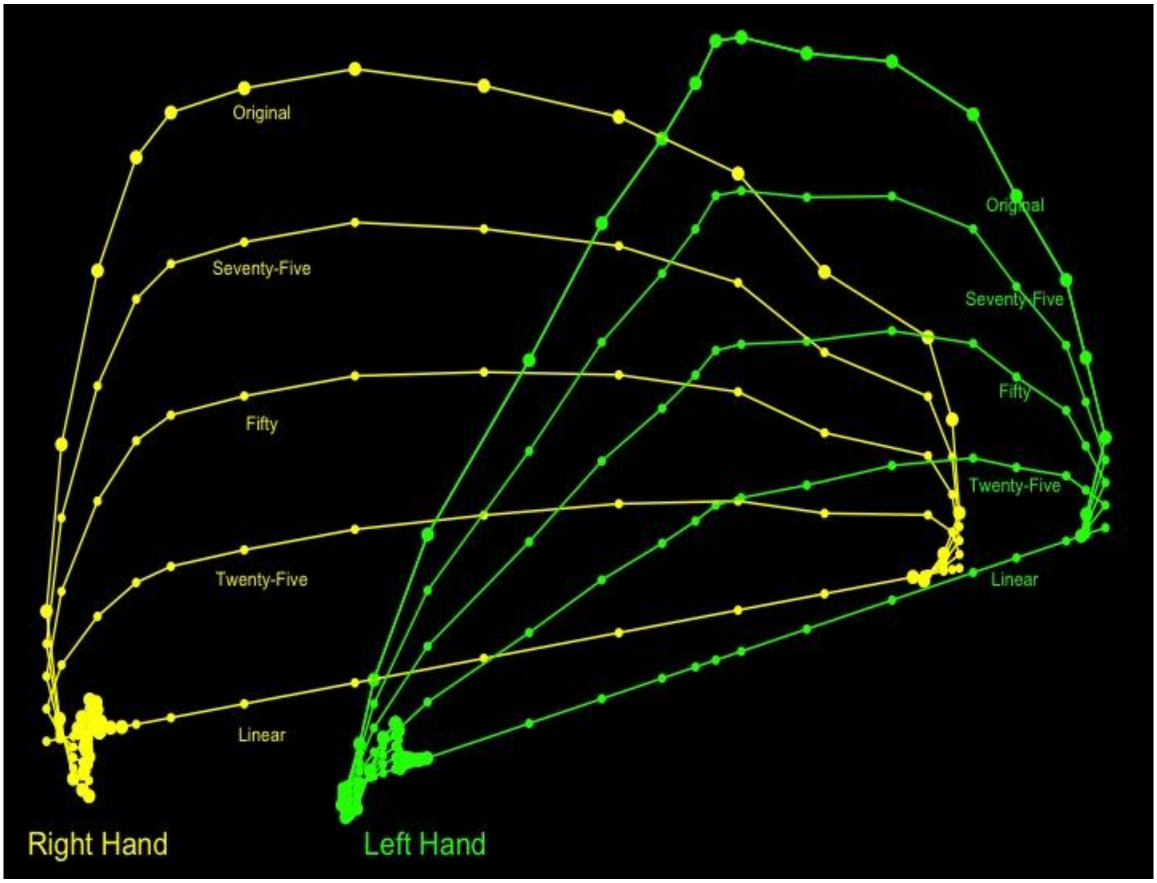

Figure 1. This picture shows how we converted the original video at 100% into the other 4 different conditions for gesture height.

We converted a video of someone explaining how to solve the problem into a ball video. We used two differently colored balls that contrast well against a black background to represent hand gestures, and manipulated the heights of the transfer gestures on increments of 25% from 0-100%. Thus, we converted one video into five ball videos for the 100%, 75%, 50%, 25%, and 0% conditions. We first annotated the original video to appropriately label all gestures as iconic, deictic, metaphoric, beat, or a combination of some of them. Then we labeled which of these gestures were transfer gestures for later use. We then recorded both positions of hand gestures frame by frame throughout the whole video. We selected the knuckle of the index finger on each hand to be the arbitrary and consistent marker of the hand position. We selected the transfer gestures that we annotated earlier and manipulated the heights parametrically relative to the original gesture height. We labeled the original height as 100%, while the manipulated heights were 75%, 50%, 25% and 0% of that original height. These manipulated heights are the five conditions for the experiment. Non-transfer gestures were not manipulated to keep the cross-modal synchrony as natural as possible.

Procedure:

We obtained consent from participants to conduct the experiment. We instructed the participants to follow the two rules in order to solve the Tower of Hanoi. They are allowed to move one disk at a time, the topmost disk on any peg can be moved to another peg, and they cannot place a larger disk on top of a smaller one, only smaller ones on top of larger ones. Then we told them that they will watch a video of someone explaining how to solve the puzzle, and then they will solve it themselves. We randomly selected participants for each condition of 100%, 75%, 50%, 25% or 0%.

After they watched one of the ball videos on the computer, we displayed the Tower of Hanoi puzzle on the screen. They solved it once, and we asked them to explain how they solved the problem. And then they solved and explained their solution again.

We tracked their mouse movements for analysis. We asked them a few questions to ensure that they did not solve this problem before recently, and they did not know the goal of this experiment.

RESULTS

Our first analysis explored the average amount of moves per condition. We wanted to see how many moves on average it took for participants to solve the Tower of Hanoi puzzle. The average number of moves per condition can be seen in Table 1. In an analysis of variance (ANOVA) with the number of moves in the first solution as the outcome variable and the condition as the predictor variable, there was no effect of condition on the number of moves (F(4,134)=0.39, ns). Thus, participants across conditions did not seem to differ in their understanding of how to solve the Tower of Hanoi.

Table 1. Average number of moves per condition

Condition	0%	25%	50%	75%	100%
Mean	31.50	37.46	35.22	36.04	37.46
Standard deviation	15.72	27.83	25.75	23.79	17.45

Our subsequent analyses explored whether variation in the gestures that were seen was associated with variation in listener behavior when solving the Tower of Hanoi. As a first look at possible effects of condition on listener behavior, we analyzed the maximum height reached by the mouse on each move of the solution. We expected that participants who saw conditions with higher gestures might produce higher mouse movements. We used a linear mixed effects model to analyze the maximum height of the mouse that participants used across the first 15 moves in their solution. Because p-values in mixed effects models with crossed random effects structure cannot be estimated, we adopted t>2 as our standard for statistical significance in this and subsequent analyses. We predicted the maximum mouse height given the condition, and the move number. We also included a random subject intercept. As shown in table 2, we found that 50% and 75% conditions have the highest curvatures, followed by 100%, and lastly, 25% and 0% looked comparable as a set. Thus, there was not a gradient effect of condition on mouse height, although participants did show evidence of being influenced by the gestures that they saw.

Table 2. Average maximum height for the first 15 moves per condition

Condition	0%	25%	50%	75%	100%
Amount of moves	198.56	202.07	222.27	219.25	213.40

Our next analysis examined the complete mouse trajectory produced when moving disks in each condition from peg to peg. We used a quadratic model to account for curvature in the mouse trajectories. X-coordinates were standardized in order to enable averaging across moves with different starting and ending points. We predicted the y-coordinate of the mouse position, given the interaction of the quadratic of the standardized x-coordinate, the condition, and all simpler effects. We also included a random intercept, main effect of x-coordinates, and quadratic effect of x-coordinates for each subject. We again restricted our analysis to mouse movements produced in the first 15 moves, the minimum required to solve the problem. In this model, there was a significant interaction between the quadratic of the x-coordinate and the experimental condition, indicating that the average mouse trajectories had different curvatures across conditions. Table 3 shows the results of this analysis, including the interaction between the quadratic of the x-coordinate and the experimental condition. We found that the 0% condition clearly has a significant mouse curvature, marked by the t value that is over 2. We also compared the other conditions to the 0% condition, which is our baseline. We find that 50% and 75% conditions are statistically significant from 0%, while 25% and 100% are not, but also, 100% is not statistically significant from either 0% or 50% and 75%. Lastly, we found that curvature of the moves become attenuated as listeners are getting closer to finishing solving the puzzle as statistically significant (β = -2.49, t = -14.95).

Table 3. Results of the model predicting participants' mouse height.

Condition	Estimated Standard deviation	Error	t-value
Intercept	218.47	6.75	32.37
25%	3.51	9.36	0.38
50%	23.71	9.44	2.51
75%	20.69	9.36	2.21
100%	14.85	9.36	1.59
Moves	-2.49	0.17	-14.95

Using the same analysis, we then compared the mouse curvatures in each condition against the 0% condition as our baseline. Table 4 shows the results of this analysis, including the interaction between the quadratic of the x-coordinate and the experimental condition. The curvature of the mouse movement in the 0% condition was negative indicating downward curvature (β = -111.03, t = -6.45). The curvature of the mouse movements in the 25% condition was more negative than that in the 0% condition, although this difference did not reach significance (β = -10.77, t = -0.44, ns). The curvature of the mouse movements in the 50% condition was significantly more negative than that in the 0 condition (β = -55.55, t = -2.26). The curvature of the mouse movements in the 75% condition was also significantly more negative than that in the 0 condition (β = -49.61, t = -2.04). The curvature of the mouse movements in the 100% condition was not significantly different from any of the conditions (β = -30.14, t = -

1.24). As can be seen in Figure 2, the 50% condition had the most curvature for mouse movement than the other conditions, followed by 75%, 100%, and then 25% and 0%. These findings indicate that the average curvature of the mouse movements produced in each condition were different from one another, although they did not show the predicted ordered effect from 100%-0%. Figure 2 demonstrates finding very well.

Table 4. Results of the model predicting participants' mouse movement trajectory.

	β	Std. Error	t-value
(Intercept)	160.62	3.13	51.39
XStan	100.60	17.62	5.71
I(XStan^2)	-111.03	17.22	-6.45
Condition25	4.78	4.42	1.08
Condition50	9.08	4.46	2.04
Condition75	10.56	4.42	2.39
Condition100	9.60	4.42	2.17
XStan:Condition25	9.62	24.92	0.39
XStan:Condition50	55.30	25.15	2.20
XStan:Condition75	46.15	24.92	1.85
XStan:Condition100	28.05	24.92	1.13
Condition25:Std Mouse X^2	-10.77	24.36	-0.44
Condition50: Std Mouse X^2	-55.55	24.58	-2.26
Condition75: Std Mouse X^2	-49.61	24.35	-2.04
Condition100: Std Mouse X^2	-30.14	24.35	-1.24

Thus, although participants were sensitive to the varying gesture heights that they observed, fine-grained and ordered sensitivity was not seen across the first 15 moves in their solution.

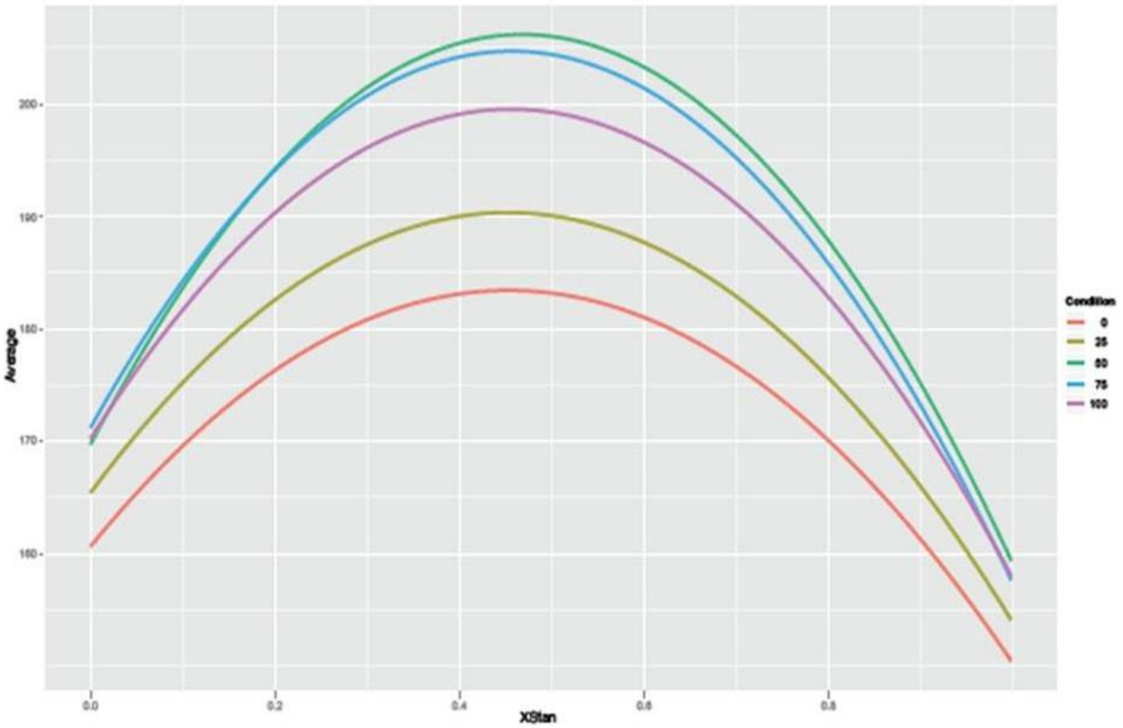

Figure 2. This graph shows the average mouse trajectory for each condition.

We expected that any effect of visual condition would be most robust at the beginning of listeners' solutions, before they have a chance to learn how to move the disks. According, in our final analysis, we used a linear mixed effects model to estimate the mouse trajectory that participants used to move the top disk from the left peg on the very first move of their solution. In this analysis, we predicted the y-coordinate of the mouse position, given the interaction of the x-coordinate and condition and all simpler effects. We also include a random intercept and a random effect of x position for each subject. We restricted our analysis to mouse movements in the leftmost quarter of the screen, because mouse movements became more curved over time. In our model, there was a significant interaction between the x-coordinate and the experimental condition, indicating that the mouse trajectories had different slopes. Table 5 shows the results of our model.

Table 5. Model predicting participants' starting trajectory of mouse movements.

	β	Std. Error	t value
(Intercept)	125.07	3.96	31.54
MouseX	0.64	0.03	20.08
Condition25	3.01	5.52	0.56
Condition50	-4.86	5.82	-0.83
Condition75	-11.08	5.34	-2.07
Condition100	-16.27	5.30	-3.07
MouseX:Condition25	-0.03	0.04	-0.57
MouseX:Condition50	0.04	0.05	0.93
MouseX:Condition75	0.09	0.04	2.09
MouseX:Condition100	0.13	0.04	3.05

We expected that the slope of the trajectory would be influenced by condition. The slope of the mouse movement in the 0% condition was positive indicating upward movement (β = 125.07, t =31.54). The slope of the mouse movements in the 25% condition was less than that in the 0 condition, although this difference did not reach significance (β = -0.03, t = -0.57, ns). The slope of the mouse movements in the 50% condition was larger, but not significantly greater than, that in the 0 condition (β = 0.04, t = 0.93, ns). The slope of the mouse movements in the 75% condition was significantly greater than that in the 0 condition (β = 0.09, t = 2.09). The slope of the mouse movements in the 100% condition was significantly greater than that in the 0 condition (β = 0.13, t = 3.05). These coefficients reveal that the slope of the mouse trajectory increased parametrically with increasing height of the ball gestures from 25% to 100%.

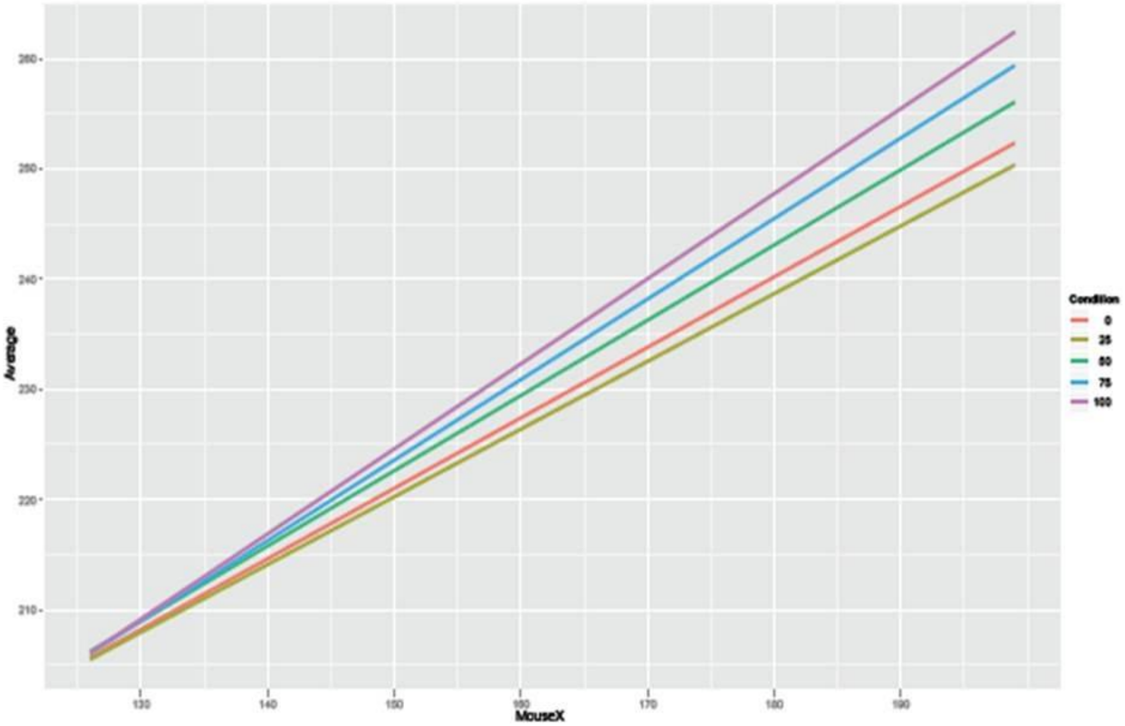

Figure 3. Starting mouse trajectories as predicted from our model.

Figure 3 depicts these results. Thus, participants were able to pick up on the continuous differences of gesture heights and reflect them accordingly in their mouse movements.

One possibility is that these gradient starting trajectories are actually the result of averaging over different distributions of upward and leftward trajectories. We then looked at the raw data of the starting trajectories in each condition. In figure 4, we see that the slopes of the starting trajectories show a wide range of values in each condition, suggesting that our results are not the result of averaging over different distributions of highly distinct trajectories.

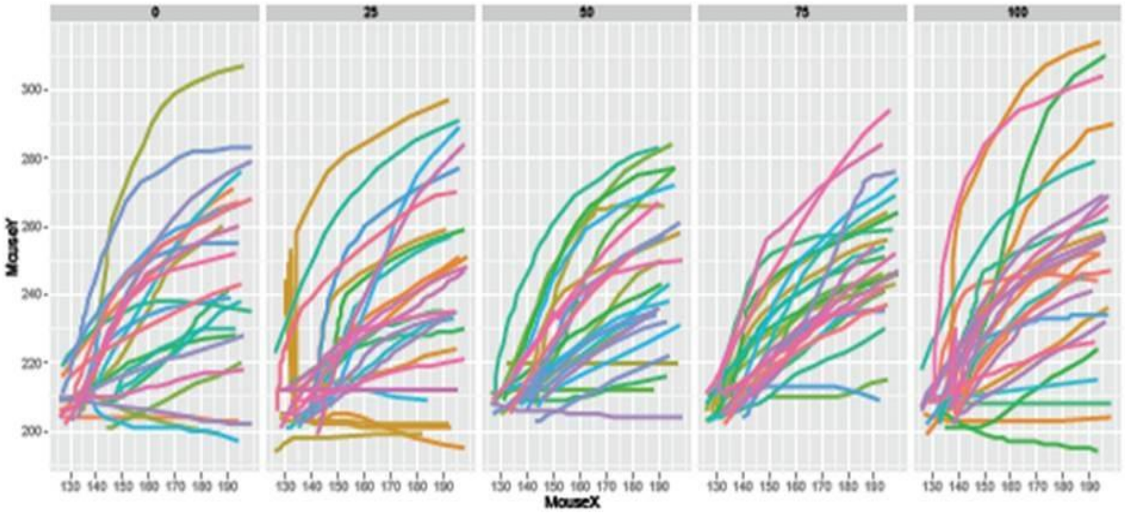

Figure 4. Raw data of the starting mouse trajectories. Each line represents the mouse trajectory of one subject in our experiment

DISCUSSION

We found that there was a graded effect of observed dot movement height on the slope of the starting trajectory. We found that listeners did not differ in the amount of moves it takes to solve the puzzle, which means that condition they were in did not affect the way they solved the puzzle. We also found that listeners were influenced by the stimuli, as there were reliable differences in both the height of movements as well as a parametric effect on the starting trajectory across conditions. We looked at the averaged curvature of the mouse trajectories for the first 15 moves for the Tower of Hanoi puzzle. We found that the 25% and 100% conditions were not statistically different from all the other conditions, but the 50% and 75% conditions, as a set, were statistically different from the 0% condition. We looked at the starting trajectory of the mouse movement and found that the trajectory was higher in for the conditions with higher gestures. Our results suggest that participants are able to differentiate between the gesture heights on a continuum and reflect that information in their subsequent movements.

We found evidence consistent with a graded effect in the initial trajectory but not in the average trajectories over the first 15 moves. The fine-grained information in gesture may not have been robust enough to lead to a sustained effect on listeners' movements. Of course, a failure to find this effect does not necessarily mean it does not exist. Even if participants were affected by the gradient information in the gestures that they observed, our measure may not have captured this influence. Although mouse movements have the potential to demonstrate gradient sensitivity, in order to explore trajectories across moves, trajectories with different starting and ending points need to be mapped into the same space. When participants pick up a disk, they pick different positions on that same disk. Second, there are different sized disks to be picked up, and the disks are located at varying heights on the peg, depending on whether the peg is previously occupied by other disks or not. Third, participants moved the disks in different paths on different moves, and different distances (either one peg away or two pegs away). All of these factors influence mouse trajectory, and a large amount of data would be necessary to account for all of these effects on movement. Thus, our measure of movement trajectory necessarily involves some distortion increasing variance and decreasing the potential sensitivity of this measure.

These results do suggest that participants are initially sensitive to fine-grained information in gesture, and that this effect may not last over time, and may even eventually become more categorical. Presumably, participants acquire their own experience with the Tower of Hanoi, and may be less influenced by their observation of someone else's experience over time.

A question that might arise is whether this effect is truly due to an underlying continuous representation. Alternatively, it is possible that the graded effect we saw in the starting trajectory is the result of averaging across underlying categorical representations. So for example, participants in each condition produced high and low gestures, and when we average them for each condition, it seems to be that the participants in all conditions produced graded gestures. We took that into account and observed the raw data of the starting trajectory for each condition. These data were consistent with a gradient effect.

There were some additional confounds in this experiment that may have affected our results. Because the transfer gestures in the stimuli varied with respect to height, when we parametrically reduced the gesture height for all gestures in an explanation, we presumably also parametrically reduced the variance in gesture height across gestures in the explanation. If participants are affected by contrast across observed gesture, we would expect smaller effects in the smaller conditions, because the difference between highly curved and less curved gestures was less.

A second problem is that the video for each condition was a mirror image of the movements that participants would produce when solving the Tower of Hanoi task, reflecting natural face-to-face dialogue with another person. Therefore, the movements that participants observed were incongruent to the movement the participants will make when they solve the puzzle. This may have weakened our effect, and explain partly why the results changed when the heights were averaged out across conditions. Participants are slower and have a harder time solving a task that is in the opposite direction of their expectation to solve the task (Richardson, Spivey, & Cheung, 2001; Wohlschläger & Wohlschläger, 1998).

More generally, the Tower of Hanoi puzzle is not an easy task to solve or to explain to someone else. Although it is a great task to use in gesture research, because it elicits a great amount of gestures (Kita & Davies, 2009), it is possible that some participants were confused or did not know what they were doing as they were solving the puzzle, and may have increased variability in our data. In addition, participants did not observe a person explaining how to solve the Tower of Hanoi task, which may have provided an underlying structure facilitating mental transformation. Instead, they saw two moving dots on a black background and the voice of someone explaining how to solve the problem. They never had previous exposure to these stimuli, and having to mentally rotate the perspective of the steps to the solution to solve the puzzle may have been cognitively taxing for them. This issue could be resolved in future studies by taking into account that the generated videos must have gesture movement in the same direction as direction the participants will solve the Tower of Hanoi. This can be done by horizontally flipping the ball videos.

In a broader sense, embodied cognition/distributed cognition demonstrates that cognitive processes are embedded in real life, physical experiences with the world. In our study, participants did not need to attend to the visual information at all. They could have learned to solve the Tower of Hanoi from the spoken instructions. However, we found that participants did attend to the visual information. Spivey et al. (2004) talked extensively about cognition as not merely an abstract processing of the world, or what he called internalism. "Rather, [the] mind appears to be an emergent property that arises among the interactions of a brain, its body, and the surrounding environment-which, interestingly, often includes other brains and bodies." In our case, participants are influenced by minimal visual stimuli. They seem to pay attention to the environment in which speech is produced in addition to the speech itself. If we were to extend this to real gestures and nonverbal representation in general, these stimuli are embedded with rich information that has the potential to influence observers' minds, and their subsequent behavior. Clearly understanding someone else is not just about understanding their abstract message, but rather includes physical aspects of the communicative context.

The present study was able to tap into how participants are able to pick up subtle information from gestures. Hanna & Brennan (2007) found that not only did listeners follow speakers' eye gaze as speakers were describing the objects, but they did so even after the speakers looked at an instruction card and tried to look for the objects they will begin to describe. Similarly, the participants in our study were able to pick up very subtle visual cues from very minimal stimuli. This indicates that people are very influenced by subtle information during communication. Cook & Tanenhaus (2009) found that listeners showed more curved mouse trajectories when they listened to the speakers who solved the physical form of the puzzle than those who solved the puzzle on the computer. This shows that listeners picked up on curved information from the gestures rather than speech. This study suggests that people are also able to view and process gestures in a gradient fashion. This is important for future studies, because this indicates that if researchers are not careful in controlling for subtle information in gestures, they may introduce unnecessary confounds in their experiments.

This present study showed that participants are influenced by the speaker's gestures, even if they are depraved stimuli. Similarly, in natural conversation, Kimbara (2006) showed that listeners mimic speakers' gestures to show that they are attentive to the conversation. Listeners also mimic to make sure that they are correctly understanding what the speaker is saying (Holler & Wilkin, 2011; Kimbara, 2006). Similarly, our present study showed that listeners mimicked the speaker's gestures in the video, though the speaker's gestures were in the form of dots moving around on the screen.

Mol et al (2012) used gesture speech match versus mismatch to test whether gestural mimicry emerges as the result of priming, or because it is conducive to communicate effectively. In comparison to the gesture speech match, where participants mimicked all gestures, participants in the gesture speech mismatch condition did not use the same gestures as the speaker in the video, presumably because they were not aligned with the semantic meaning of the speech. Therefore, apparent mimicry in gesture or body language is not just for the sake of copying another person, but to establish common ground subconsciously to facilitate social interaction, or common ground for understanding information in speech. This further supports the fact that the listeners in our study did not just mimic the speaker's gestures to solve the puzzle, but because these gestures were related to the puzzle solution.

Gestures can be used in a variety of ways depending on the context in which they are used. They help the speaker speak their mind, or find words when they are dealing with tip of tongue phenomenon (Krauss, 1998; Kita, 2000). From the speaker perspective, speakers with greater spatial resources produce more non-redundant gesture-speech combinations than other speakers, which help them communicate more effectively (Hostetter & Alibali, 2011). On the other side, gestures also help the listener understand where the speaker is coming from, and help aid in communication (Kendon, 1994). Gestures also help the listener glean information that is otherwise not found in speech (Beattie & Shevolton, 2006; Cook & Tanenhaus, 2009; MacGregor, 2003; Cassell et al., 1999; Goldin-Meadow & Sandhofer, 1999). Some of the information that can be found in gestures that is not found in speech can be the speaker's ability to fluidly shift perspective in conversation to help the listener understand a narrative story (Cassell et al., 1999), or information about what children are trying to say (Goldin-Meadow &

Sandhofer, 1999), especially if they are at the toddler stage. Not only that, gestures help people remember speech better (Church et al., 2007; Woodall & Folger, 1985). These are the same results that Quinn-Allen (1999) found when teaching French to non-French speakers using gestures. Gestures help in communication and in learning.

More directly related to this present study, it is important for listeners to be able to pick up pertinent information from gesture that is not found in speech. Imagine a scenario where a person is helping another with directions to a certain destination. Sometimes, a speaker makes a mistake in speech while giving directions, but had the correct information in their gesture. If the listener only attended to the speech, then surely they will get lost trying to find their destination. Moreover, in more complicated social scenarios, sometimes an individual's intent is ambiguous (Kelly et al., 1999; Obermeier et al., 2011). For example, an individual is indirectly requesting something, uses sarcasm, or states passive-aggressive remarks to another individual. If the listener fails to pick up on the double meaning of these remarks by not attending to gestures or body language in general, over time, they risk miscommunication, which can potentially bring harm to themselves or break down the relationship with the speaker, because they fail to pay attention and the speaker can take this as a sign of disrespect and lack of care or love. Gestures are conducive in understanding ambiguous meaning in speech, and it is essential to pick up on subtle information from gestures if listeners want to be effective communicators and social beings.

When people have a conversation, they use gestures. As we have shown, listeners are sensitive to information in gesture. Moreover, they appear to be sensitive to fine-grained detail – even small changes in the form of a gesture have reliable effects on listeners' behavior. We listen with our eyes as well as our ears.

BIBLIOGRAPHY

Alibali, M. W., Evans, J. L., Hostetter, A. B., Ryan, K., & Mainela-Arnold, E. (2009). Gesture-speech integration in narrative: Are children less redundant than adults? *Gesture, 9*, 290-311.

Allen , L. Q. (1995). The effects of emblematic gestures on the development and access of mental representations of French expressions. *Modern Language Journal, 79*, 521-9.

Beattie, G. & Shovelton, H. (2006). When size really matters: How a single semantic feature is represented in the speech and gesture modalities. *Gesture, 6*, 63-84

Beilock S. L. & Goldin-Meadow S. (2010). Gesture changes thought by grounding it in action. *Psychological Science. 21*, 1605-1610

Cassell, J., McNeill, D., & McCullough, K.-E. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics and Cognition, 7,* 1-34.

Church, R. B., Garber, P., & Rogalski, K. (2007). The role of gesture in memory and social communication. *Gesture*, 7(2), 137-158

Cook, S. W., & Tanenhaus, M. K. (2009). Embodied communication: speakers' gestures affect listeners' actions. *Cognition, 113*, 98-104.

Cook, S.W. (in preparation). Listeners integrate a variety of visual information with speech.

Dale , R.; Kehoe, C. & Spivey , M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition, 35,* 15-28.

Emmorey, K., McCullough, S., & Brentari, D. (2003). Categorical perception in American Sign Language. *Language & Cognitive Processes, 18,* 21-46.

Galati, A. & Brennan, S.E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62,* 35-51.

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

Goldin-Meadow , S. Sandhofer , C. M. (1999). Gesture conveys substantive information about a child's thoughts to ordinary listeners. *Developmental Science, 2*, 67-74.

Habets, B., Kita, S., Shao, Z., Özyürek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. *Journal of Cognitive Neuroscience, 23,* 1845-1854.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language, 57,* 596-615.

Hary, J. M. & Massaro, D. W. (1982). Categorical results do not imply categorical perception. *Perception & Psychophysics, 32(5)*, 409-418.

Holler, J.; Stevens, R. (2007). The Effect of Common Ground on how Speakers use Gesture and Speech to Represent Size Information. *Journal of Language and Social Psychology*, *26*, 4-27.

Holler, J. & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes, 24,* 267-289

Holler, J. & Wilkin, K. (2011). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behavior, 35(2),* 133-153.

Hostetter, A. B. & Alibali, M. W. (2011). Cognitive skills and gesture-speech redundancy: Formulation difficulty or communicative strategy? *Gesture, 11(1),* 40-60.

Kelly, S. D., Barr, D., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: the role of speech and gesture in comprehension and memory. *Journal of Memory & Language, 40,* 577-92.

Kendon, A. (1994). Do gestures communicate? A review. *Research on Language and Social Interaction, 27,* 175-200.

Kimbara, I. (2006). On gestural mimicry. *Gesture, 6(1),* 39-61.

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

Kita, S. & Davies, T. S. (2009). Competing conceptual representations trigger co-speech representational gestures. *Language and Cognitive Processes, 24*, 761-775.

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

Langton, S. R. H., & Bruce, V. (2000). You must see the point: Automatic processing of cues to the direction of social attention. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 747-757.

Liberman, A. M., Harris, K. S. Hoffman, H., & Griffith , B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358-368.

MacGregor, D. (2004). Real Space Blends in Spoken Language. *Gesture, 4*, 75-89.

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

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2012). Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language, 66(1),* 249-264.

Obermeier, C., Holle, H., & Gunter, T. C. (in press). What iconic gesture fragments reveal about gesture-speech integration: When synchrony is lost, memory can help. *Journal of Cognitive Neuroscience, 23*, 1-16.

Pastore, R. E., Ahroon, W. A., Puleo J. S., Crimmins D. B., Golowner L., & Berger, R. S. (1977). Common-factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance, 3(4),* 686-696.

Pisoni, D. B. & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America, 55(2),* 328-333.

Richardson, D.C., Spivey, M.J., & Cheung, J. (2001). Motor representations in memory and mental models: Embodiment in cognition. *Proceedings of the Twenty-third Annual Meeting of the Cognitive Science Society*, (pp.867-872), Erlbaum: Mawhah, NJ.

Runeson, S., & Frykholm, G. (1981). Visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception & Performance, 7*, 733-740

Spivey, M. J.; Grosjean, M.; Knoblich, G. (2005). Continuous attraction toward phonological competitors. PNAS Proceedings of the National Academy of Sciences of the United States of America, 102, 10393-10398.

Spivey, M. J., Richardson, D. C., Fitneva, S. A. (2004). Thinking outside the Brain: Spatial Indices to Visual and Linguistic Information. Henderson, John M. (Ed); Ferreira, Fernanda (Ed), *The interface of language, vision, and action: Eye movements and the visual world*, (pp. 161-189). New York, NY, US: Psychology Press.

Thompson, L. A. & Massaro, D. W. (1986). Evaluation and integration of speech and pointing gestures during referential understanding. *Journal of Experimental Child Psychology, 42(1),* 144-168.

Treffner, P., Peter, M., & Kleidon, M. (2008). Gestures and phases: The dynamics of speech-hand communication. *Ecological Psychology, 20*, 32-64.

Wohlschläger , A. Wohlschläger , A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 397-412.

Woodall, W. & Folger , J. (1985). Nonverbal cue context and episodic memory: On the availability and endurance of nonverbal behaviors as retrieval cues. *Communication Monographs, 52,* 319-333.