ORIGINAL PAPER

Using and Seeing Co-speech Gesture in a Spatial Task

Alexandra Suppes · Christina Y. Tzeng · Laura Galguera

Abstract Visibility of speakers' conversational hand gestures has been shown to facilitate listeners' comprehension. Is this true for all types of gestures? In Experiment 1, speakers were videotaped describing apartment layouts from memory. When describing complex compared to simple layouts, speakers used more representational gestures. In particular, they used more iconic–deictic gestures, or hand movements that represent *both* an object or action, *and* direction or location. This suggests that the frequency of complex representational gestures increases as a function of task difficulty. To assess the extent to which these gestures facilitate speech comprehension, in Experiment 2 a new group of participants either watched the videos of the apartment descriptions (audio + video condition) or only heard the descriptions (audio-only condition) and drew each corresponding layout. Although drawing accuracy did not differ as a function of condition, the more iconic–deictic gestures produced during the apartment descriptions, the less accurate listeners' drawings were. Together, findings from Experiments 1 and 2 suggest that iconic–deictic gestures reflect task difficulty for the speaker and do not necessarily facilitate comprehension for the listener.

A. Suppes (✉)
Department of Psychology, Columbia University, 402 Schermerhorn Hall, New York, NY 10027, USA
e-mail: suppes@gmail.com; acs2008@columbia.edu

C. Y. Tzeng
Department of Psychology, Emory University, 36 Eagle Row, Room 270, Atlanta, GA 30322, USA

L. Galguera
Department of Quantitative Economics at the Faculty of Economics and Business Studies, University of Oviedo, Avda del Cristo S/N, 33006 Oviedo, Asturias, Spain

Introduction

Speakers often produce co-speech gestures, or unplanned, fluent hand movements while talking. Previous research suggests that co-speech gestures aid in the communicative process, but whom these gestures help—the speaker during production, or the listener during perception, or both—has been much debated. One view claims that gestures convey information that augments the information in the speech signal and thus enhances listener comprehension (e.g., Beattie and Shovelton 1999; Kendon 1983). In a meta-analysis of 63 studies examining the effect of gesture on speech comprehension, Hostetter (2011) concluded that there was a modest but significant effect of seeing co-speech gestures on message comprehension when those gestures were (1) spatial or motor related or (2) non-redundant with co-speech language. Based on these results, Hostetter claims that "Gestures do benefit comprehension, and this benefit is independent of any benefits gestures may have for a speaker's production" (p. 311).

An alternative view is that gestures serve little communicative function for the listener but rather aid the speaker in formulating messages (e.g., Krauss 1998). In particular, co-speech gestures have been found to aid both lexical retrieval (Butterworth & Hadar 1989; Hadar et al. 2001; Krauss 1998; Pine et al. 2007; Rauscher et al. 1996) and the conceptual planning of utterances (Melinger and Kita 2007; Hostetter et al. 2007; Alibali et al. 2000). In addition to helping the structure of speech, gestures also facilitate the maintenance of spatial representations in working memory (Wesp et al. 2001) such that individuals gesture more often when descriptions feature spatial components (for a review, see Hostetter 2011). According to this view, the primary function of co-speech gestures is to enhance speech fluency.

Interestingly, not all gesture types facilitate descriptions of spatial information. Spatial descriptions tend to feature representational gestures, which comprise of both *iconic gestures*, which resemble the gesture's referent in form, and *deictic gestures*, which indicate a location or path (McNeill 1992). These representational gestures reflect spatial thinking and are more likely to co-occur when producing spatial words than when producing non-spatial words (Alibali 2005; Beattie and Shovelton 1999; Krauss 1998). Representational gestures also aid lexical retrieval when linguistic information has visuo-spatial components (Krauss and Hadar 1999; McNeill 1992) and enhance a speaker's spatial memory (Morsella and Krauss 2004).

The utility of representational gestures also increases with task difficulty such that when visuo-spatial information is more complex, speakers' production of representational gestures increases. Hostetter et al. (2007), for example, asked participants to describe the shape of dot arrays (e.g., "The top three dots form a triangle"). When arrays lacked guiding lines that simplified the task, speakers produced more representational gestures. These and similar findings provide evidence for the information packaging hypothesis (Kita 2000), which claims that gestures help speakers organize complex, visuo-spatial information into packages, or units, suitable for speaking. Speakers produce more representational gestures when the visuo-spatial information cannot be easily verbalized. Representational gestures can thus be conceptualized as a mode of thinking that helps translate complex visuo-spatial information into linguistic output.

Goldin-Meadow and Alibali (2013) have further suggested that representational gestures are, at least in part, intentionally used by speakers to help communicate salient information. There is circumstantial evidence for this. First, speakers gesture more when they can see one another than when they cannot (Bavelas et al. 2008; Mol et al. 2011). Second, speakers do not use some types of gestures when alone. Specifically, when

interlocutors cannot see one another, the frequency of beat gestures, which convey no semantic information but have a rhythmic relationship to the accompanying speech, remains high, while the frequency of representational gestures decreases (Alibali et al. 2001; Clark and Krych 2004).

Although it is important to consider how representational gestures may be intentionally used to communicate information, speaker intentionality does not imply that greater frequency of their use achieves that end. Considering that representational gestures are produced more when describing complex visuo-spatial information (Hostetter et al. 2007), it is possible that the presence of these gestures may reflect accompanying speech that is uninformative or disfluent.

Current Studies

In the current studies we address the utility of co-speech hand gestures, with a particular focus on the role of specific types of gestures during the production and comprehension of messages. We asked the following three questions: (1) What specific types of gestures do speakers produce in spatial tasks of varying complexity (Experiment 1)?, (2) To what extent do the different types of gestures produced in Experiment 1 facilitate listener comprehension (Experiment 2)?, and (3) To what extent does it matter if listeners see the gestures that accompany the spoken message?. In Experiment 1, speakers described simple and complex visuo-spatial stimuli (apartment layouts), and in Experiment 2, participants either watched (video + audio condition) or listened (audio-only condition) to those descriptions and attempted to draw the respective layouts.

What types of gestures will speakers produce when describing complex versus simple apartments in Experiment 1? Given previous evidence that speakers produce more representational gestures when the visuo-spatial information cannot be easily verbalized (e.g., Hostetter et al. 2007), we predicted that speakers should produce more representational gestures when describing complex relative to simple apartments.

Experiment 1

Method

Participants

Sixteen female native English speakers (Barnard University undergraduates between 18 and 22-years-old) participated in Experiment 1 for course credit in our research lab. Participants were recruited using web-based sign-up procedures approved by the institutional review board at Columbia University. One speaker was removed from our analysis because she did not gesture once (Experiment 1, $N = 15$). For clarity, participants in Experiment 1 will be referred to as "speakers."

Materials

Apartment layouts served as stimuli (for examples, see "Appendix"). Apartments were drawn digitally on 9 × 10 grids and shared particular features, including at least one

bedroom and bathroom, a front door, one kitchen, and one living room. Some apartment layouts were manipulated to be more complex by increasing the number of bedrooms, bathrooms, and hallways. In a pilot study, lab members attempted to describe the layouts and noted which they found most complex. This generated apartment layouts that we considered "complex" and layouts that we considered "simple."

Procedure

After providing informed consent to participate in our laboratory, speakers sat in an armless chair facing a digital camcorder that recorded their movements and wore a head-mounted microphone to record their speech. Speakers were told that they would describe three apartment layouts in English so that future participants would be able to recreate the apartments by following their descriptions. Instructions to speakers were ambiguous about whether future participants would view the videos and/or only hear the audio component of their descriptions.[1]

Speakers were given one layout at a time and had unlimited time to memorize each one. When speakers finished studying the given layout, they handed it to the experimenter who sat behind them. Speakers were instructed to describe the apartment from memory, directing their speech to the camcorder. This procedure was repeated for each of the layouts, with the order of describing simple and complex apartments first counterbalanced across participants.
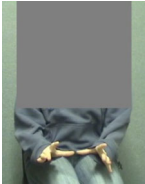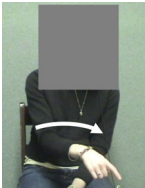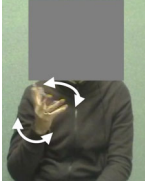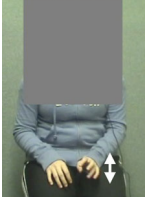
Coding of Gestures

Transcription and gesture coding was completed by a native English speaker (CYT). During gesture coding, the videos were played at both original and 30-frame-per-second speeds using Final Cut Express video editing software.

Coding of the gestures occurred in three passes (Duncan 2002; McNeill 1992). First, apartment descriptions were transcribed and segmented into short utterances reflecting the grammatical structuring of the speech. Most segments consisted of a verb and its associated modifiers (e.g., "If you walk in through the front door"). Second, gesture onset and offset were identified based on changes in hand movement, shape, and position. Gesture onsets and offsets occurred both within and across word boundaries (McNeill 1992) and were noted as such in the transcription. Gestures were identified as all hand movements produced, excluding those that were irrelevant to the task (e.g., scratching one's arm). Finally, each gesture was labeled as a particular type using McNeill's classification system (1992; for definitions and examples, see Table 1). In this pass, we identified (1) iconic, (2) deictic, (3) metaphoric, and (4) beat gestures. Because many of our speakers produced a single gesture type that represented *both* iconic and a deictic information, we identified an additional gesture type, (5) combined iconic–deictic. To establish reliability of our coding scheme, a second coder completed the final pass by labeling each gesture in a random sample of apartment descriptions. To measure the agreement in gesture labels between the two coders, we calculated the Cohen's Kappa coefficient (Cohen 1960), which takes the

---

[1] The study was designed to additionally explore differences in first and second language production (English and Spanish), though this is not the focus of this paper: For the purpose of this paper, we will only describe results of the apartments that were described in English. Order of language (first or second) was counterbalanced across participants. A native English speaker gave all instructions in English.

Table 1 Definitions and samples of gestures produced in Experiment 1

Gesture type	Definition	Example	Description and accompanying speech
Iconic	Hand shape or movement that resembles the referent, usually an action or object, as if the referent is embodied in the hand		Speaker cups both hands to form a rectangular shape; "…[the apartment is again square]"
Deictic	Pointing movements that indicate direction or location, create relative locations in gesture space between concrete or imaginary objects or people		Speaker uses her right index finger to point left to the imaginary location of a room relative to other rooms in an apartment; "…[it's right across from the kitchen]"
Iconic–deictic	Hand shape or movement that represents *both* an object or action, *and* direction or location.		Speaker moves her left hand across her body to represent both the act of walking and the direction of movement "…[if you walk in through the front door]"
Metaphoric	Hand forms or movements representing abstract concepts, such as size comparison or approximation		Speaker rotates her right palm in a back and forth motion to represent the concept of *a little bit;* "…the living room [is fairly small]"
Beats	Formless hand movements that convey no semantic information but have a rhythmic relationship to the accompanying speech		Speaker taps left index finger up and down left in pace with syllables; "…[will] be [Bedroom] [One]"

percentage agreement between two coders based on chance into account (Kappa = 0.83 for gesture identification).

Statistical Approach

Did speakers use more of a specific type of gesture when describing simple compared to complex apartments? In the MIXED procedure of SAS (SAS Institute 2001), we constructed a series of multilevel models (one for each type of gesture) to predict how frequently speakers used a given gesture type when describing apartments that were either

complex or simple, controlling for the total word count of that description. Multilevel modeling allowed us to take full advantage of our repeated design, so we estimated intercepts that were unique to each speaker, effectively accounting for each speaker's unique frequency of gesturing. As coefficient estimates are unstandardized in multilevel modeling, we provide estimates, labeled b, their standard errors, as well as the 95 % confidence interval for each estimate.

Results and Discussion

Did speakers use more of a specific type of gesture when describing simple compared to complex apartments? We did not see a significant difference in the speakers' use of iconic gestures, $F(1, 14) = 0.12$, $p = 0.74$, deictic gestures, $F(1, 14) = 0.05$, $p = 0.82$, metaphoric gestures, $F(1, 14) = 0.08$, $p = 0.78$, or beat gestures, $F(1, 14) = 0.14$, $p = 0.71$, during descriptions of simple versus complex layouts.

Although manipulating apartment layout complexity did not affect speakers' use of iconic, deictic, metaphoric, or beat gestures, speakers' use of the combined iconic–deictic gesture did vary by the complexity of the layout they described. As shown in Table 3, speakers used more iconic–deictic gestures when describing a complex versus a simple apartment, $b = 2.29(0.71)$, $F(1, 14) = 10.47$, $p < 0.006$, suggesting that speakers used, on average, 2.29 more iconic–deictic gestures when describing a complex versus simple apartment (95 % CI; 0.76, 3.83).

In addition to estimating the fixed effect of apartment complexity on gesture rate (such that speakers used more iconic–deictic gestures when describing complex apartments), we assessed between-speaker variance in gesture rate. The second-to-last line of Table 3, labeled "Speaker Intercept" shows an estimation of the random effect of the speaker. Results show significant heterogeneity in speakers' use of the iconic–deictic gesture, $z = 0.2.38$, $p < 0.037$ suggesting that the unique intercepts for each speaker were significantly different from one another. Put another way, in addition to using more iconic–deictic gestures when describing complex apartments, speakers showed significant variance in their use of iconic–deictic gestures.

Our original prediction that speakers would use more representational gestures when describing complex apartments was partially confirmed, given that speakers used more of the combined iconic–deictic gesture, but not other types of representational gestures. Why might this have been the case? Although we did not analyze the content of speakers' verbal output, it is possible that while describing a more complex apartment layout, speakers may be more inclined to use semantically complex motion phrases (e.g., "and if you turn left"), leading them to generate iconic–deictic gestures, which include *both* an object or action, *and* direction or location. This interpretation would lend support to the information packaging hypothesis, which states that the production of representational gestures helps speakers organize visuo-spatial information and occurs more often with descriptions of increasing complexity (Kita 2000). Alternatively, the increase in iconic–deictic gestures may reflect the greater spatial memory demand on the part of the speaker when describing complex apartments. This interpretation would replicate previous work that manipulates the complexity of visuo-spatial information (e.g., Hostetter et al. 2007). Unfortunately, our current data are not designed to test whether increase in the iconic–deictic gesture was due to increased demand on speakers' verbal output, memory, or both. However, regardless of the cause of increased iconic–deictic gestures, our data do suggest that higher numbers of iconic–deictic gestures may reflect difficulty on the part of the speaker to produce a fluent description. Experiment 1 on its own does not address to what extent these gestures are

useful in message comprehension. In Experiment 2, we examine how variation in gesture use affects the accuracy with which listeners are able to draw apartments that match the layouts speakers described.

Experiment 2

We originally predicted that speakers in Experiment 1 would produce more representational gestures when describing complex apartment layouts compared to simple ones, but found this only to be true for the combined iconic–deictic gesture. We will therefore modify our predictions for Experiment 2 to reflect the use of iconic–deictic gestures, rather than the more general category of representational gestures. If the frequency of iconic–deictic gesture production increased when describing complex apartment layouts because the speaker found those layouts more challenging to describe (as predicted by the information packaging hypothesis), or because they found them harder to remember (as suggested by Hostetter et al. 2007), we would expect that speakers who used more iconic–deictic gestures made descriptions that were harder to follow. Therefore we predict that the higher frequency of iconic–deictic gestures should impact message comprehension regardless of whether listeners saw the descriptions or only heard them.

Method

Participants

One hundred fifty-eight native speakers of English (Columbia University undergraduates, 70 % female, M age years = 20.9, SD = 5.3) participated in Experiment 2 for course credit. Participants were recruited using web-based sign-up procedures approved by the institutional review board at Columbia University. Participants in Experiment 2 will be referred to as "listeners."

Stimuli

The 45 apartment descriptions in English from 15 speakers generated in Experiment 1 served as stimuli. Video recordings were converted in QuickTime and audio files were converted in Sound Studio. Listeners were presented either the audio + video files or the audio-only files.

Procedure

Listeners were randomly assigned to the audio + video condition (N = 72) or the audio-only condition (N = 86). Listeners viewed or heard each apartment description twice and were given a clipboard that contained blank grids to draw the apartment layouts described. To provide listeners with a starting point, the location of the front door was placed on the blank layout. The descriptions were presented to listeners over Sennheiser HD 280 Pro Circumaural headphones connected to 14-in. display Macintosh G3 computers. Each listener was randomly presented six apartment layouts: three in English and three in Spanish. For the purpose of this paper, we will only describe data from the three layouts listeners were presented in English.

Table 2 Descriptive statistics for number of gestures used (Experiment 1) and listener accuracy (Experiment 2) by complexity of apartment layout

	Simple layouts		Complex layouts	
	Mean	SD	Mean	SD
Experiment 1-speaker's gesture rate				
Iconic	5.00	4.83	5.23	5.57
Deictic	4.27	5.15	4.81	4.27
Beats	3.60	3.16	4.19	4.69
Metaphoric	1.47	2.33	1.68	2.17
Iconic–deictic	1.80	2.40	4.26	3.60
Total gestures	16.13	13.63	20.16	16.06
Word count	125.27	65.80	141.32	81.47
Experiment 2-listener's drawing accuracy				
Accuracy (%)	77.86	19.23	67.85	20.30

Layout Accuracy

Listeners' drawings generated in Experiment 2 were scored for accuracy relative to the original layout that speakers were asked to describe in Experiment 1. Points were awarded to each drawing based on several objective criteria: (1) Did the drawing include all rooms in the original layout?, (2) Were the rooms placed in the correct location?, and (3) Were the relative sizes of each room correct? There were two additional criteria that were used to deduct points: (1) Did the listener rotate the layout (e.g., such that all rooms on the right appeared on the left)?, and 2) Did the listener not use the full grid area? Because the number of rooms differed across complex and simple layouts, percent correct was calculated for each drawing based on the particular layout it was meant to depict. Summarized in the second half of Table 2, there was a range in listeners' accuracy scores, with drawings from simple layouts earning higher percentage scores than drawings from complex layouts ($M = 77.9 \%$; $SD = 19.2 \%$ and $M = 67.9 \%$; $SD = 20.3 \%$).

To establish reliability of our map accuracy-coding scheme, a second coder followed the same coding in a random sample of apartment layouts. Because the rating system required few subjective judgments, we had very high internal consistency between the two coders (Cronbach's $\alpha = 0.95$).

Statistical Approach

Six multilevel models were built to predict the accuracy of apartment drawings with change in accuracy expressed by change in percentage points. The first model examined how apartment accuracy was affected by (1) the condition of the listener (audio + video vs. audio-only) and (2) if a complex or simple apartment was being described when (3) controlling for the word count of a description. We capitalized on the repeated observations design of our study and allowed variability due to speakers and listeners to serve as random effects, allowing us to estimate means that are unique to each speaker and each listener.

In our second series of multilevel models we had two goals: first, to determine if listeners drew more accurate apartment layouts when the speaker used more or fewer gestures (e.g., Did listeners make worse drawings when speakers used more iconic–deictic gestures?), and second, did it matter if listeners saw those gestures (e.g., Was there a

Table 3 Number of iconic–deictic gestures used, by apartment complexity, in Experiment 1

Effect	Fixed effects				95 % Confidence limits	
	Estimate	SE	t Value	p	Lower	Upper
Intercept	1.95	0.78	2.49	<0.026	0.27	3.63
Word count[a]	0.014	0.0076	1.84	<0.088	−0.0024	0.030
Apt layout complexity[b]	2.29	0.71	3.24	<0.01	0.76	3.83
	Random effects				95 % Confidence limits	
	Estimate	SE	z Value	p	Lower	Upper
Speaker intercept	4.24	2.38	2.38	0.037	1.80	19.38
Residual	4.88	1.28	3.81	<0.01	3.10	8.81

$df = 14$

[a] Mean centered

[b] Apt layout complexity is coded, simple = 0; complex = 1

significant interaction of a speaker's number of iconic–deictic gestures and the condition of the listener?). Building on the structure of our first model, we built a series of five models (one for each gesture type). Each unique model predicted the effect of the number of times a speaker used (4) iconic, (5) deictic, (6) metaphoric, (7) beat, and (8) iconic–deictic gestures on the accuracy of subsequent listeners' apartment drawings.

In these models, we also examined our second research goal: to assess whether being able to see specific gesture types helped listeners draw more accurate apartments. Consequently, for each of the five models, we added (9) an interaction term for number of times a speaker used a given gesture by the condition of the listener (e.g., In the model testing the effect of iconic–deictic gestures, we had an interaction term for the number of iconic–deictic gestures a speaker used by the condition of the listener).[2]

Finally, given that we are interested in the unique effect of a given gesture type and not a speaker's overall rate of gesturing, we controlled for (10) the frequency of all other gestures the speaker used during that description. Because we were interested in the accuracy between original layout and drawing, we did not control for the accuracy of the specific description listeners heard or saw. All models were built using the MIXED procedure of SAS (SAS Institute 2001). Below, we report unstandardized coefficient estimates of the chance in percentage points (b), their standard error, and the 95 % confidence interval for these estimates.

Results and Discussion

Did listeners draw more accurate apartment layouts when they could see the speakers' descriptions? As summarized in Table 4, our first multilevel model revealed that the accuracy of layout drawings were comparable for listeners in the audio + video and audio-only conditions, $F(1, 154) = 1.04$, $p = 0.31$, suggesting that visual access alone did not help listeners draw more accurate layouts. Unsurprisingly, drawings of simple apartments were scored as more accurate than complex apartments $b = -10.42$ (1.96), $F(1,$

---

[2] We are grateful for this suggestion by an anonymous reviewer of our manuscript.

Table 4 Listener's apartment drawing accuracy (from 0 to 100), as predicted by the condition of the listener, in Experiment 2

Effect	Fixed effects				95 % Confidence limits	
	Estimate	SE	t Value	p	Lower	Upper
Intercept	77.29	1.95	39.69	<0.01	73.11	81.46
Word count[a]	0.016	0.017	0.95	0.34	−0.017	0.049
Apt layout complexity[b]	−10.42	1.96	−5.33	<0.01	−14.28	−6.56
Listener's condition[c]	−2.22	2.18	−1.02	0.31	−6.52	2.07

	Random effects				95 % Confidence limits	
	Estimate	SE	z Value	p	Lower	Upper
Speaker intercept	12.69	10.22	1.24	0.11	4.12	166.87
Listener intercept	70.65	23.39	3.02	<0.01	40.48	153.50
Residual	321.10	27.54	11.66	<0.01	273.29	382.72

$df = 154$

[a] Mean centered

[b] Apt layout complexity is coded, simple = 0; complex = 1

[c] Condition is coded, audio + video = −0.5; audio-only = 0.5

154) = 28.39, $p < 0.001$, suggesting that drawings of simple apartment layouts were, on average, 10 % points higher than drawings of complex layouts, (95 % CI; −14.28, −6.56).

Random effects of apartment layout accuracy were tested to assess between-speaker and between-listener variance. Looking at Table 4, under "Speaker Intercept," we can see that there was little heterogeneity between speakers, $z = 1.24$, $p < 0.11$. Put another way, our sample of 15 speakers did not contain participants who produced "better" or "worse" descriptions across all the layouts described: Each speaker had descriptions that allowed listeners to draw accurate (and inaccurate) apartment layouts.

Where we did see heterogeneity was in the random effect of listeners. Our model estimated a unique intercept for each listener to begin predicting the accuracy of his or her apartment drawings. In Table 4, under "Listener Intercept," we see a great deal of heterogeneity between listeners, $z = 0.3.02$, $p < 0.0012$, meaning that unique listener's intercepts were significantly different from one another. Put another way, regardless of what descriptions they heard or saw, some listeners simply drew more accurate layouts compared to others. The pattern of between-speaker and between-listener heterogeneity will be the same for all models run in Experiment 2, and so we will not discuss them further.

Although listeners in the audio + video condition did not draw more accurate apartments, it is possible that the *number* of gestures produced during descriptions influenced the accuracy of apartment drawings. In our second series of multilevel models, we tested if the number of times a speaker used a given gesture affected the accuracy of listeners' drawings. Summarized in Table 5, we found a negative association between the number of iconic–deictic gestures used in a description and the accuracy of apartment drawings listeners made, $b = −0.78(0.39)$, $F(1, 151) = 4.10$, $p < 0.045$. When speakers used more iconic–deictic gestures, listeners drew worse apartments (95 % CI −1.54; −0.019) or, for each additional iconic–deictic gesture speakers used, listeners' accuracy decreased only by 0.78 % points. Nonetheless, the negative effect of speakers' use of iconic–deictic gestures

Table 5 Listener's apartment drawing accuracy (from 0 to 100), as predicted by talker's use of iconic deictic gestures

Effect	Fixed effects				95 % S Confidence limits	
	Estimate	SE	*t* Value	*p*	Lower	Upper
Intercept	76.03	2.03	37.51	<0.01	71.68	80.38
Word count[a]	−0.0077	0.022	−0.35	0.73	−0.052	0.036
Apt layout complexity[b]	−8.55	2.13	−4.02	<0.01	−12.75	−4.35
Listener's condition[c]	−2.20	2.17	−1.01	0.31	−6.49	2.09
# of other gestures[a]	0.31	0.12	2.48	0.014	0.062	0.55
# of iconic–deictic gestures[a]	−0.78	0.39	−2.02	0.045	−1.54	−0.019
# of iconic–deictic gestures[a] X Listener's condition[c]	0.28	0.53	0.52	0.60	−0.78	1.33

	Random effects				95 % S confidence limits	
	Estimate	SE	*z* Value	*p*	Lower	Upper
Speaker intercept	12.69	11.48	1.11	0.13	3.74	283.26
Listener intercept	71.81	23.25	3.09	<0.01	41.57	152.90
Residual	315.43	27.27	11.57	<0.01	268.13	376.52

$df = 151$

[a] Mean centered

[b] Apt Layout Complexity is coded, simple = 0; complex = 1

[c] Condition is coded, audio + video = −0.5; audio-only = 0.5

becomes meaningful if we consider this result in context: Imagine we compare two speakers, one whose gesture rate is one standard deviation above versus one standard deviation below the mean rate of iconic–deictic gesture use when describing a complex apartment (shown in Table 2, $M = 4.26 \pm 3.60$). The speaker one SD above the mean would make just under eight iconic–deictic gestures, compared to the speaker one SD below the mean who would make < 1 iconic–deictic gesture. Given that each iconic–deictic gesture a speaker used predicts a decrease in drawing accuracy of 0.78 % points, this seven-gesture difference would translate to a difference of five percentage points in accuracy (e.g., a change, say, from 80 % accurate to 85 % accurate).

This series of models also addressed our second research goal: to assess if it mattered whether or not listeners saw speakers' particular gestures. Specifically, was the interaction between a speaker's iconic–deictic gesture use and a listener's condition significant? Table 5 reveals speakers' use of iconic–deictic gestures influenced listeners regardless of whether or not the listener saw those iconic–deictic gestures: the interaction term of the speaker's use of iconic–deictic gestures and the listener's condition was not significant $F(1, 151) = 0.27$, $p = 0.60$. Taken together with our findings from Experiment 1, that speakers used more iconic–deictic gestures when they are describing a complex apartment, these results suggest that speakers use iconic–deictic gestures when either visuo-spatial memory demands or when speaking demands are high.

Our second series of multilevel models aimed to assess the influence of each specific gesture type on apartment-drawing accuracy also revealed a positive association between

Table 6 Listener's apartment drawing accuracy (from 0 to 100), as predicted by talker's use of beat gestures

Effect	Fixed effects				95 % Confidence limits	
	Estimate	SE	*t* Value	*p*	Lower	Upper
Intercept	77.36	2.00	38.69	<0.01	73.07	81.65
Word count[a]	0.018	0.025	0.71	0.48	−0.033	0.069
Apt layout complexity[b]	−10.56	1.97	−5.37	<0.01	−14.45	−6.68
# Other gestures[a]	−0.086	0.17	−0.51	0.61	−0.42	0.25
Listener's condition[c]	−2.20	2.17	−1.01	0.31	−6.49	2.09
# of beat gestures[a]	0.65	0.30	2.18	0.03	0.060	1.23
# of beat gesture[a] X listener's condition[c]	0.12	0.43	0.27	0.79	−0.74	0.97

	Random effects				95 % Confidence limits	
	Estimate	SE	*z* Value	*p*	Lower	Upper
Speaker intercept	15.51	12.65	1.23	0.110	4.98	214.73
Listener intercept	70.63	23.34	3.03	<0.01	40.49	153.20
Residual	318.21	27.54	11.55	<0.01	270.45	379.93

$df = 151$

[a] Mean centered

[b] Apt layout complexity is coded, simple = 0; complex = 1

[c] Condition is coded, audio + video = −0.5; audio-only = 0.5

the number of beat gestures a speaker made during an apartment description and the accuracy of subsequent apartment. As summarized in Table 6, listeners drew more accurate apartment layouts when exposed to descriptions which featured more beat gestures, $b = 0.65(0.30)$, $F(1, 151) = 4.75$, $p < 0.031$. This association suggests that speakers who used more beat gestures when making their descriptions provided listeners with a description that enabled them to make more accurate apartment drawings (95 % CI, 0.062; 1.24). As with the number of iconic–deictic gestures, this change in percentage points is quite small, only 0.65 percentage points. If we again consider a speaker who was one SD above versus one SD below the mean in use of beat gestures when describing a complex apartment (from Table 2, $M = 4.19 \pm 4.69$), we are looking at a difference of over eight beat gestures that would lend itself to a difference in predicted accuracy score of over five percentage points. As with the negative effect of iconic–deictic gestures, we found that the positive effect of speakers' use of beat gestures was not influenced by the condition of the speaker: The interaction term of speaker's use of beats by listener condition was not significant $F(1, 151) = 0.08$, $p = 0.77$, meaning it didn't matter if listeners could see these beat gestures or if they only heard the description with which they were associated.

Our second series of models revealed that no other gesture type was associated with the accuracy of apartment drawings: speakers' use of iconic, $F(1, 151) = 0.59$, $p = 0.44$, deictic $F(1, 151) = 1.16$, $p = 0.28$, or metaphoric, $F(1, 151) = 0.06$, $p = 0.81$ gestures did not affect the accuracy of apartment drawings. As with the a speaker's use of iconic–deictic and beat gestures, the interaction of a listener's condition and speaker's frequency

of iconic, $F(1, 151) = 0.10$, $p = 0.75$, or deictic $F(1, 151) = 0.11$, $p = 0.75$, or meta-phoric $F(1, 151) = 0.56$, $p = 0.45$ gesture was not significant.[3]

General Discussion

In Experiment 1, we explored differences in how speakers use various types of gesture when describing simple and complex apartment layouts. Results suggest that the more complex an apartment layout, the higher the number of iconic–deictic gestures speakers used. In Experiment 2, we asked listeners to draw apartment layouts based on the descriptions generated in Experiment 1. Being able to see a video of the apartment descriptions did not improve listeners' accuracy of subsequent drawings. Interestingly, regardless of whether a simple or complex apartment was being described and regardless of whether the listeners could see the speaker's gestures, descriptions that featured more iconic–deictic gestures led to less accurate apartment drawings.

Why does gesture production during an apartment description affect the accuracy of apartments drawn from that description? Our findings suggest that the use of iconic–deictic gestures reflects task difficulty such that they were used more when layouts were complex (Experiment 1). Descriptions featuring more iconic–deictic gestures resulted in less accurate apartment layout drawings (Experiment 2). Interestingly, it was the presence of these gestures in the description and not whether the listeners saw them that affected the accuracy of subsequent drawings. This leads us to speculate about why speakers use iconic–deictic gestures. First, the finding that speakers produced more iconic–deictic gestures when describing complex apartments aligns with the information packaging hypothesis (Kita 2000). According to this view, representational gestures aid in the planning of a verbal message. Specifically, representational gestures occur most frequently when speakers organize spatial information into units suitable for verbal expression. Given the difficulty of organizing the relative location, size, and type of rooms in each complex apartment layout, speakers might have produced more iconic–deictic gestures to facilitate verbal descriptions of these features. Second, even if they were used to help spatial recall in this task, as previous work would suggest (De Ruiter 1998; Morsella and Krauss 2004; Wesp et al. 2001), these findings do not inform the extent to which iconic–deictic gestures were *successfully* recruited to aid spatial memory in our task. Studies where hand mobility is restricted suggest that had participants not been able to produce these iconic–deictic gestures, their descriptions would have been even less successful (Morsella and Krauss 2004). In contrast with theories implying that gesture facilitates spatial memory (de Ruiter 1998; Morsella & Krauss 2004; Wesp et al. 2001), the information packaging hypothesis makes explicit claims regarding the role of gesture in lessening speaking rather than memory demands. Although the present data cannot fully disentangle these two accounts, they do support the view that representational gestures, specifically iconic–deictic gestures, reflect task difficulty level.

There is an alternative explanation for why higher frequency of iconic–deictic gestures in descriptions predicted worse apartment drawings. Perhaps, rather than the

---

[3] The model summarized in Table 5 reveals that the overall use of speakers' other gestures during a given description, after accounting for the use of iconic-deictic gestures, was positive, $b = 0.31(0.12)$, $F(1, 151) = 6.13$, $p < 0.02$ (95 % CI 0.062; 0.55). This positive association in overall gesture rate is driven by the positive association between beat gestures and apartment drawing accuracy revealed in Table 6. As can be seen in Table 6, when accounting for beat gestures, the association between total gestures used and accuracy is not significant.

production of iconic–deictic gestures reflecting poor memory of an apartment layout, the difference in layout drawing performance lies in the decoding of what these iconic–deictic gestures often represented: the description of motion. By this account, it may be easier to recreate an apartment layout from a description that simply states where the rooms are located than from a description in which the speaker is navigating in an imagined space.

Although the focus of this work is the role of representational hand gestures, we were surprised to find that apartment layout descriptions that featured a higher frequency of beat gestures resulted in better drawings. Unlike representational gestures, beat gestures do not convey any semantic content (McNeill 1992). Although the role of beat gestures is not well understood (Casasanto 2013), there is some evidence that beats play a role in speech prosody and may aid in subsequent comprehension. Speech segments that are accompanied by a beat gesture are acoustically differentiated. For example, Krahmer and Swerts (2007) found that words that occurred along with beat gestures had high-frequency formants, thereby enhancing the acoustic prominence of those words for the listener. Taken together with our findings from Experiment 2, it is possible that beat gestures either help speakers produce comprehensible speech or at least are a reflection of the speaker's communicative clarity. Not mutually exclusive from this view is the possibility that the production of beat gestures reflects semantic fluency (Hostetter and Alibali 2007). Given claims that manual movements that have no direct relation to the semantic content in accompanying speech can activate speech-related brain regions and facilitate lexical access (Ravizza 2003), beat gestures may operate in a similar manner. The link between beat gestures and shifts in the acoustical features or semantic fluency of the apartment descriptions might explain why beat gestures did not need to be seen in Experiment 2 to have a positive impact on subsequent apartment drawings.

Finally, there are limitations to our approach. In the audio-only condition, we removed listeners' access to any visual information about the speaker, including his or her mouth movements, which listeners benefit from being able to see (Jesse & Massaro 2010). Why did our video condition listeners not benefit from being able to see speakers' mouth movements? Perhaps the video recordings, aimed to show speakers' head and torso, did not have clear enough images of the mouth. Nonetheless, being able to see the gestures or the mouth movements did not facilitate comprehension on the part of the listeners.

The role of co-speech gesture may vary by situational context and is affected by the highly social nature of communication. Speakers constantly adapt how they speak and to whom they are speaking. This has been shown at a linguistic level (Pickering and Garrod 2004), at a paralinguistic level (Pardo 2006), and in gesture. For example, the use of representational and beat gestures increases when mothers speak to language-learning infants (Iverson et al. 1999), when foreign language teachers speak to language learners (Allen 2000), and when interlocutors have less common ground (Jacobs & Garnham 2007). Moreover, speakers adapt their gestures to match the idiosyncrasies of the person with whom they are communicating, dyadically synchronizing their gestures during a conversation (Bavelas et al. 1988; Kimbara 2006; Wallbott 1995). Some authors (e.g., Goldin-Meadow 2003) have even suggested that gestures may play a role in promoting positive affect between speakers. Further still, negotiation researchers have found that, compared to phone negotiations, those who interact face-to-face are more likely to behave cooperatively (Purdy et al. 2000) and to build dyadic rapport (Drolet and Morris 2000) than those who negotiate over the phone. Taken together, these findings emphasize the importance of

considering the nature of the communicative environment when interpreting the role of gesture.

The present findings directly compare the communicative effects of different types of gestures, implying that even within the category of representational gestures, the role of gesture varies as a function of gesture type. Although gestures may play an important social role in communication, our findings suggest that the production of representational gestures does not necessarily facilitate message comprehension, particularly when the presence of those representational gestures reflects high spatial memory or speaking demands.

Appendix

See Fig. 1.
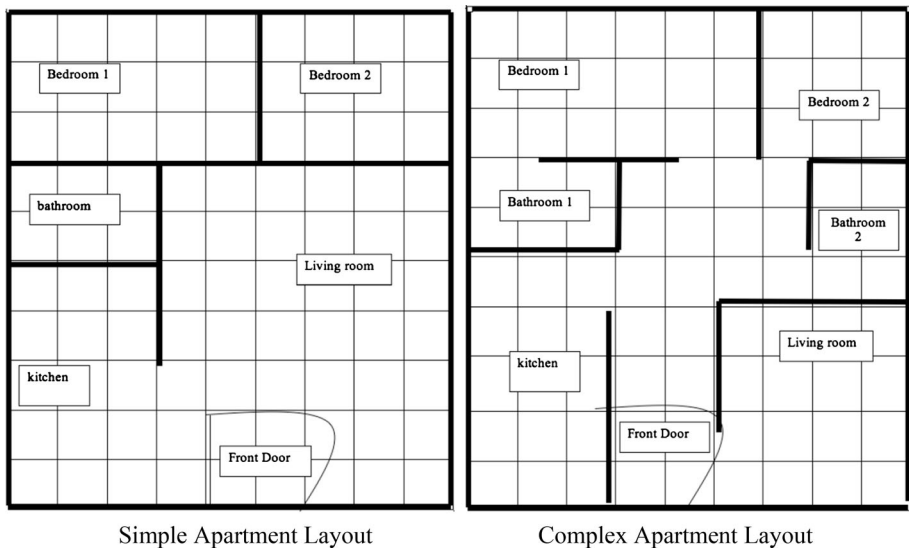
Fig. 1 Sample of a simple and a complex apartment layout

References

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating and thinking about spatial information. *Spatial Cognition and Computation, 5*, 307–331. doi:10.1207/s15427633scc0504_2.

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44*, 169–188. doi:10.1006/jmla.2000.2752.

Alibali, M. W., Kita, S., & Young, A. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes, 15*, 593–613. doi:10.1080/016909600750040571.

Allen, L. Q. (2000). Nonverbal accommodations in foreign language teacher talk. *Applied Language Learning, 11*, 155–176.

Bavelas, J. B., Black, B., Chovil, N., Lemery, C. R., & Mullett, J. (1988). Form and function in motor mimicry. Topographic evidence that the primary function is communicative. *Human Communication Research, 14*, 275–299. doi:10.1111/j.1468-2958.1988.tb00158.x.

Bavelas, J. B., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*, 495–520. doi:10.1016/j.jml. 2007.02.004.

Beattie, G. W., & Shovelton, H. K. (1999). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica, 123*, 1–2. doi:10. 1515/semi.1999.123.1-2.1.

Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review, 96*, 168–174. doi:10.1037/0033-295X.96.1.168.

Casasanto, D. (2013). Gesture and language processing. In H. Pashler (Ed.), *Encyclopedia of the mind* (Vol. 7, pp. 373–375). Thousand Oaks, CA: Sage Publications.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*, 62–81. doi:10.1016/j.jml.2003.08.004.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and psychological measurement. *Educational and Psychological Measurement, 20*, 37–46. doi:10.1177/001316446002000104.

De Ruiter, J. P. (1998). *Gesture and speech production*. Doctoral dissertation at Catholic University of Nijmegen, Netherlands.

Drolet, A. L., & Morris, M. W. (2000). Rapport in conflict resolution: Accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts. *Journal of Experimental Social Psychology, 36*, 26–50. doi:10.1006/jesp.1999.1395.

Duncan, S. (2002). Gesture, verb aspect, and the nature of iconic imagery in natural discourse. *Gesture, 2*, 183–206. doi:10.1075/gest.2.2.04dun.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: The Belknap Press.

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology, 64*, 257–283. doi:10.1146/annurev-psych-113011-143802.

Hadar, U., Dar, R., & Teitelman, A. (2001). Gesture during speech in first and second language: Implications for lexical retrieval. *Gesture, 1*, 151–165. doi:10.1075/gest.1.2.04had.

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological Bulletin, 137*, 297–315. doi:10.1037/a0022128.

Hostetter, A. B., & Alibali, M. W. (2007). Raise your hand if you're spatial: Relations between verbal and spatial skills and gesture production. *Gesture, 7*, 73–95. doi:10.1075/gest.7.1.05hos.

Hostetter, A. B., Alibali, M. W., & Kita, S. (2007). I see it in my hand's eye: Representational gestures are sensitive to conceptual demands. *Language and Cognitive Processes, 22*, 313–336. doi:10.1080/ 01690960600632812.

Iverson, J., Capirici, O., Longobardi, E., & Caselli, M. (1999). Gesturing in mother–child interactions. *Cognitive Development, 14*, 57–75. doi:10.1016/S0885-2014(99)80018-5.

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 56*, 291–303. doi:10.1016/j.jml.2006.07.011.

Jesse, A., & Massaro, D. W. (2010). Seeing a singer helps comprehension of the song's lyrics. *Psychonomic Bulletin & Review, 17*, 323–328. doi:10.3758/PBR.17.3.323.

Kendon, A. (1983). Gesture and speech: How they interact. In J. M. Weimann & R. P. Harrison (Eds.), *Nonverbal interaction*. Beverly Hills, CA: Sage Publications.

Kimbara, I. (2006). On gestural mimicry. *Gesture, 6*, 39–61. doi:10.1075/gest.6.1.03kim.

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

Krahmer, E., & Swerts, M. (2007). Effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language, 57*, 396–414. doi:10. 1016/j.jml.2007.06.005.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7*, 54–59. doi:10.1111/1467-8721.ep13175642.

Krauss, R. M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. In L. S. Messing & R. Campbell (Eds.), *Gesture, speech, and sign* (pp. 93–116). Oxford, UK: Oxford University Press.

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

Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes, 22*, 473–500. doi:10.1080/01690960600696916.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2011). Seeing and being seen: The effects on gesture production. *Journal of Computer Mediated Communications, 17*, 77–100. doi:10.1111/j.1083-6101.2011.01558.x.

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *American Journal of Psychology, 117*, 411–424. doi:10.2307/4149008.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119*, 2382–2393. doi:10.1121/1.2178720.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169–226. doi:10.1017/S0140525X04000056.

Pine, K. J., Bird, H., & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science, 10*, 747–754. doi:10.1111/j.1467-7687.2007.00610.x.

Purdy, J. M., Ney, P., & Balakrishnan, P. V. (2000). The impact of communication media on negotiation outcomes. *International Journal of Conflict Management, 11*, 162–187. doi:10.1108/eb022839.

Rauscher, F. B., Krauss, R. M., & Chen, Y. (1996). Gesture, speech and lexical access: The role of lexical movements in speech production. *Psychological Science, 7*, 226–231. doi:10.1111/j.1467-9280.1996.tb00364.x.

Ravizza, S. (2003). Movement and lexical access: Do noniconic gestures aid in retrieval? *Psychonomic Bulletin & Review, 10*, 610–615. doi:10.3758/BF03196522.

SAS Institute. (2001). *SAS system for Microsoft Windows*. Cary, NC: SAS Institute.

Wallbott, H. G. (1995). Congruence, contagion, and motor mimicry: Mutualities in nonverbal exchange. In J. Markova, C. F. Graumann, & K. Foppa (Eds.), *Mutualities in dialogue* (pp. 82–98). Cambridge, UK: Cambridge University Press.

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *American Journal of Psychology, 114*, 591–600. doi:10.2307/1423612.