Psychological Science OnlineFirst, published on September 17, 2015 as doi:10.1177/0956797615597671

Research Article

Psychological Science
1–11

DOI: 10.1177/0956797615597671
pss.sagepub.com

Iconic Gestures Facilitate Discourse Comprehension in Individuals With Superior Immediate Memory for Body Configurations

Ying Choon Wu[1] and Seana Coulson[2]
[1]Institute for Neural Computation and [2]Department of Cognitive Science, University of California, San Diego

Abstract

To understand a speaker's gestures, people may draw on kinesthetic working memory (KWM)—a system for temporarily remembering body movements. The present study explored whether sensitivity to gesture meaning was related to differences in KWM capacity. KWM was evaluated through sequences of novel movements that participants viewed and reproduced with their own bodies. Gesture sensitivity was assessed through a priming paradigm. Participants judged whether multimodal utterances containing congruent, incongruent, or no gestures were related to subsequent picture probes depicting the referents of those utterances. Individuals with low KWM were primarily inhibited by incongruent speech-gesture primes, whereas those with high KWM showed facilitation—that is, they were able to identify picture probes more quickly when preceded by congruent speech and gestures than by speech alone. Group differences were most apparent for discourse with weakly congruent speech and gestures. Overall, speech-gesture congruency effects were positively correlated with KWM abilities, which may help listeners match spatial properties of gestures to concepts evoked by speech.

Keywords

action perception, embodiment, gesture comprehension, grounded cognition, kinesthetic memory, motor resonance, multimodal discourse, working memory, haptic memory

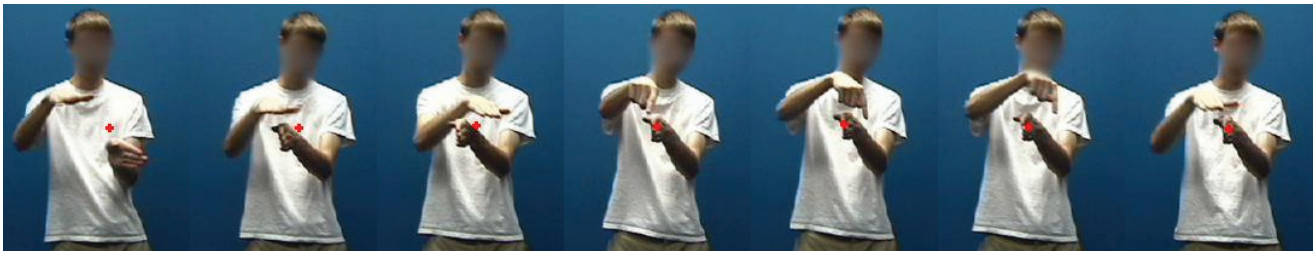
Received 8/28/14; Revision accepted 7/3/15

Iconic gestures, or gestures that depict visuospatial properties of their referents, can influence discourse comprehension. They affect on-line processing at both the word and message levels (Wu & Coulson, 2007b, 2010a) and can lead to an enhanced understanding of the visuospatial properties of discourse referents (Wu & Coulson, 2007a). Interaction-focused researchers have suggested that people understand gestures by virtue of shared bodily experience in the physical world (Goodwin, 2003; Streeck, 2002, 2008). Accordingly, we hypothesized that sensitivity to gesture meaning is linked to the listener's ability to remember the spatio-motoric particulars of the speaker's gestures.

Memory for the kinesthetic properties of gesture is important because both speech and gestures unfold dynamically (Fig. 1; see also Video S1 in the Supplemental Material available online). Information in a speaker's gestures may precede the information in the speech stream that is relevant for their interpretation. Because multimodal-discourse comprehension requires the integration of asynchronously presented information, here we examined the extent to which this process recruits memory systems for maintaining both verbal and gestural input.

Researchers have proposed that the psychological construct of working memory (WM) mediates the temporary maintenance and manipulation of information. Originally, it was thought to include a *central executive*, along with two modality-specific slave systems—the *phonological loop* and the *visuospatial sketchpad*, responsible for the rehearsal of verbal and visuospatial information,

Corresponding Author:
Seana Coulson, University of California, San Diego, Cognitive Science, 9500 Gilman Dr., Mail Code 0515, La Jolla, CA 92093-0515
E-mail: coulson@cogsci.ucsd.edu



"the countertop goes over a bit—kind of curved in the middle"

Fig. 1. Example screen shots from a multimodal-discourse video clip used in the experiment (see Video S1 in the Supplemental Material available online for the entire video clip). Stimuli were created by video-recording a speaker describing everyday activities, events, and objects while producing accompanying gestures. In this example, the concept of a curved countertop edge evolves over the course of the speaker's utterance. Consecutive gestures accompanying the speech depict the relationship between the countertop and the cabinet faces (first three frames from left to right), the curved overhang (next three frames), and the flat surface (final frame). A fixation cross was presented throughout the clip.

respectively (Baddeley & Hitch, 1974). Modern conceptions of the visuospatial sketchpad include, at the very least, dissociable visual and spatial components (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999; Klauer & Zhao, 2004), with further fractionation of the visual buffer into color and shape information, as well as a peripheral storage component for *haptic memory*, divided into kinesthetic and tactile information (Baddeley, 2012). Dual-task (Seemüller, Fiehler, & Rösler, 2011; Smyth, Pearson, & Pendleton, 1988; Smyth & Pendleton, 1989, 1990) and delayed-match-to-sample (Posner, 1967; Seemüller, Müller, & Rösler, 2012) studies support the existence of such a peripheral storage component for kinesthetic phenomena, including body configurations and patterns of movement.

If speech-gesture integration relies on temporary memory for a speaker's movements, then one might expect a positive relationship between individual differences in body-centered kinesthetic WM (KWM) abilities and sensitivity to gestures in discourse. This *spatio-motoric hypothesis* gives rise to the prediction that greater sensitivity to gestures would be associated with better KWM. To test this prediction, we measured healthy adults' capacity to remember and reproduce sequences of body movements using the movement span task (Wu & Coulson, 2014). Next, we tested whether individuals best able to temporarily remember body configurations were also those most influenced by the semantic content of iconic gestures.

Short audio files of spontaneous discourse were paired with three types of video accompaniments: (a) *congruent gestures*, which originally co-occurred with the spoken discourse; (b) *incongruent gestures*, which were simply formerly congruent videos now paired with audio such that speech and gestures no longer matched in meaning; and (c) *no gestures*—freeze frames of the speaker extracted at points without gesture. Discourse primes were followed by either related picture probes depicting a scene or an object described by the speaker or unrelated probes. Participants judged whether each picture probe was related to the preceding video clip.

Whereas the immediate, on-line effects of speech-gesture integration on lexical processing have received considerable attention (Habets, Kita, Shao, Özyurek, & Hagoort, 2011; Özyurek, Willems, Kita, & Hagoort, 2007; Wu & Coulson, 2010a, 2010b), the present approach affords a slightly more downstream, message-level view focused on conceptualization of speaker meaning as it relates to participants' KWM capacity. We expected a positive relationship between movement-span scores and the facilitative effect of iconic gestures on picture-probe classification—but no relationship between movement-span scores and the inhibitory effect of incongruent gestures. A positive relationship between KWM capacity and sensitivity to gestures would suggest that speech-gesture integration recruits KWM for ongoing analysis of spatio-motoric features of the speaker's movements—likely for the purpose of establishing mappings to conceptual knowledge about discourse referents.

Method

Participants

On the basis of existing studies of individual differences in the processing of representational gestures (e.g., Hostetter & Alibali, 2007), we recruited 90 healthy undergraduates at University of California, San Diego (52 females, 38 males). All gave informed consent and received course credit for participation. English was the primary language of all participants. Six participants were ultimately excluded because of low accuracy on the picture-relatedness task.

Picture-relatedness task

Stimuli and procedure. Discourse primes for the picture-relatedness task were constructed from continuous video footage of a speaker describing everyday activities, events, and objects to an off-camera interlocutor (Fig. 1). The speaker was told only that he should provide detailed

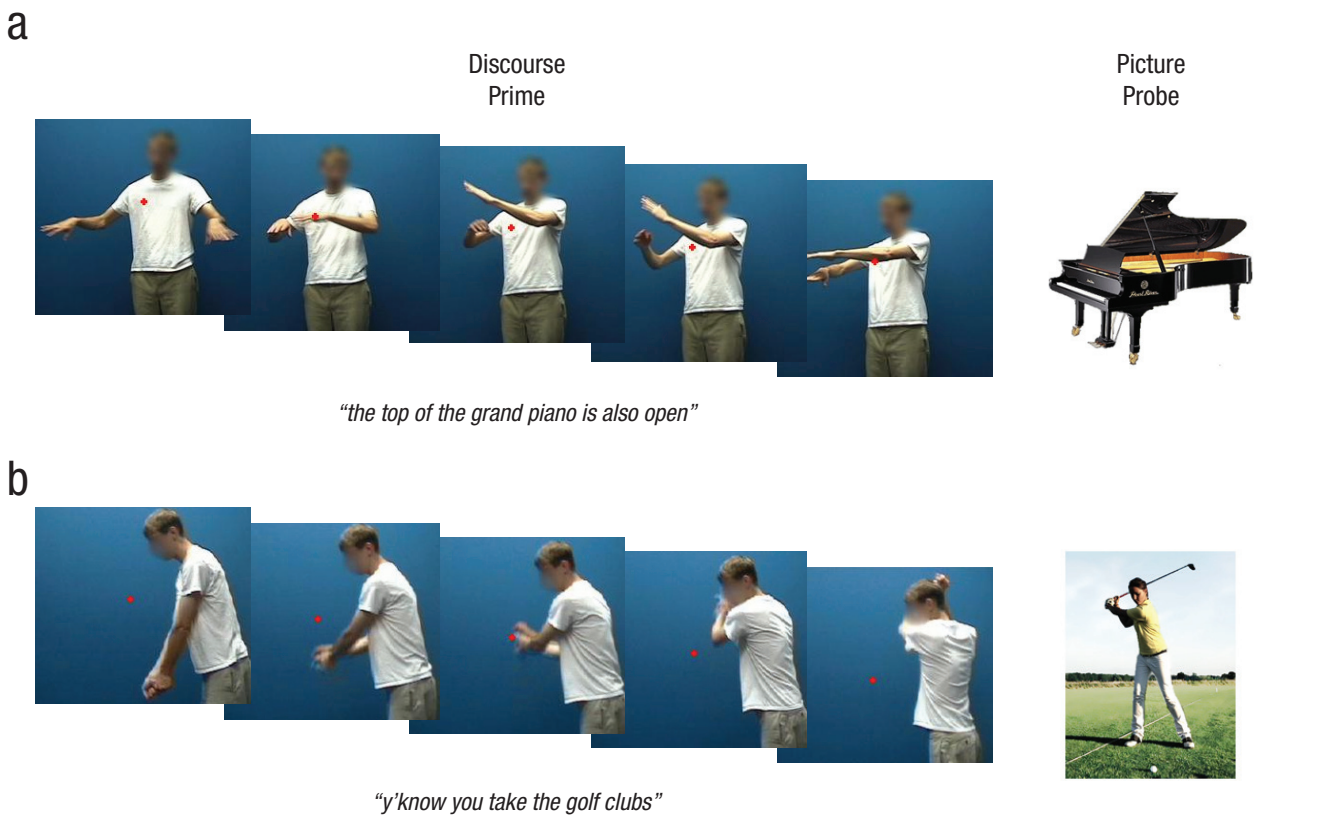


Fig. 2. Examples of audiovisual discourse primes and associated picture probes. In the discourse primes, the speaker made gestures that usually added information that went above and beyond what he conveyed verbally. In (a), the speaker's gesture indicates the orientation of the piano (see Video S2 in the Supplemental Material for the congruent version of the full clip). In (b), he gesturally depicts swinging a golf club (see Video S3 in the Supplemental Material for the congruent version of the full clip).

descriptions; he was given no guidance about whether to produce gestures and was otherwise naive to the experimenter's true motivation for filming. Short clips (2–8 s) containing both speech and iconic gestures were extracted from the video. Discourse typically centered on the use of tools and common objects (e.g., Fig. 2b; see also Video S3 in the Supplemental Material) or visual descriptions of people or objects (e.g., Fig. 2a; see also Video S2 in the Supplemental Material). In some cases, the speaker produced and held a single gesture over the course of the clip (as in Fig. 2b and the congruent gesture depicted in Fig. 3) or reiterated the same gesture several times. In other cases, he produced a series of gestures, each building off of its predecessor (as in Figs. 1 and 2a).

In many clips, gesture onset coincided with the onset of the video, and gesture offset did not occur until the end of the clip. Gesture onset was defined as the first frame in which the speaker's hands deviated from the "ready position" he assumed between gestures, either with his arms at his sides or his hands folded in front of him. Average gesture onset in experimental clips was 150 ms (*SD* = 346) from the beginning of the clip. Gesture offset was defined as the frame in which the speaker resumed the ready position; it occurred on average 225 ms (*SD* = 477) from the end of the clip. Because natural, spontaneous discourse can be difficult to comprehend when presented in such brief snippets, written titles were created to precede each video and offer supportive context.

In congruent primes, the audio and video portions of each clip were paired in their original form; this preserved the semantic congruence between the speaker's speech and gestures (examples of congruent discourse primes can be found in Videos S1 through S3 in the Supplemental Material). To form the incongruent primes, we rearranged the audio and video elements of the congruent stimuli so that the semantic information conveyed by the gestures no longer matched that in the speech (examples of incongruent discourse primes can be found in Videos S4 through S6 in the Supplemental Material). For neutral trials, video segments were replaced with freeze frames extracted from portions of the discourse stream when the speaker was not gesturing. Neutral stills were combined with the exact same audio segments used in the congruent and incongruent conditions and were presented for the same duration as their video counterparts.

Because the speaker's orofacial movements did not match the speech output in the incongruent stimuli, the

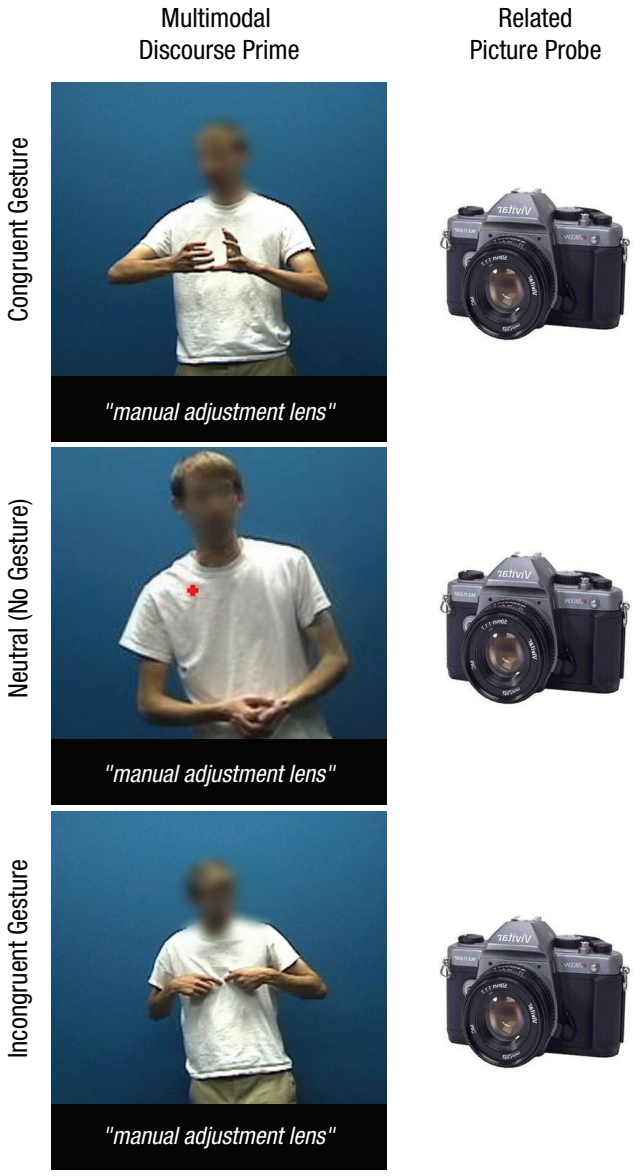


Fig. 3. Sample frame and related picture probe from each of the three congruency conditions. In each condition, participants heard clips of the same audio describing activities, events, or objects, but different video played depending on the condition. On congruent trials, video of the speaker gesturing was presented in its original form. On incongruent trials, video was taken from a different segment of the recording. On neutral trials, freeze frames were taken from parts of the video in which the speaker did not gesture. A picture probe followed each prime, and participants judged whether the picture was related to the preceding audiovisual clip.

speaker's face was blurred in all three types of discourse primes. In a separate norming study conducted with new participants drawn from the same pool as in the main experiment (N = 10), participants were asked to rate the degree of correspondence between speech and gestures on a 5-point Likert scale (5 = *highly congruent*, 1 = *highly incongruent*). Congruent videos received an average rating of 3.8 (SD = 0.8), while the average rating for incongruent videos was 2.2 (SD = 0.7).

Each congruent prime was matched with a related photographic picture probe that agreed with both the speaker's speech and gestures. Counterpart incongruent and neutral primes were paired with the same picture probes, although they were related only to the speech portion of the discourse prime. Thus, a single-factor design was employed with one type of probe (related) and three types of discourse primes (congruent, incongruent, and neutral; Fig. 3). To create a balanced quantity of trials that elicited "yes" and "no" responses, we constructed unrelated fillers by replacing the picture probes across the three prime types with images that no longer agreed with either the speech or the gestures.

Each participant viewed an equal number (n = 70) of related and unrelated picture probes, each preceded by an equivalent number of congruent, incongruent, and neutral discourse primes. The use of six randomized lists ensured that no items were repeated on any list. Across lists, however, each audio clip was presented as part of a congruent, an incongruent, and a neutral discourse prime. Probes were also distributed such that each related picture served as its own control, being paired with the congruent, incongruent, and neutral versions of the same discourse prime across lists.

Trials began with a title describing the topic of the upcoming discourse segment. Subsequently, the video (or freeze frame) was presented at a rate of 30 ms per frame in the center of a computer monitor. After the video, there was a 50-ms pause, then the picture probe appeared in the center of the screen for 500 ms, followed by a blank screen. Between 1.1 and 6.8 s (Mdn = 3.1) elapsed from the onset of the speaker's gestures to the onset of the picture probe. After the participant's response was registered, the next trial was initiated, beginning with a blank screen for 1 s.

Participants were informed that they would be watching a series of short videos in which a man describes various things. They were instructed to read each title silently to themselves and then to watch and listen to each video (or simply to listen when the video was a freeze frame) while holding their index fingers on the "A" and "L" keys. When the picture probe appeared, they were to press the "A" or "L," depending on whether the picture was related or unrelated to what preceded. Response keys were counterbalanced across participants.

Analysis. Response latencies were computed from picture-probe onset to the time of key press. Correct responses to related probes were analyzed if they occurred within 2.5 standard deviations of the participant's mean response time (i.e., picture-classification time) for each condition. Approximately 4% of trials were

trimmed. Picture-classification times, aligned both by subjects and items, and the percentage of accurate responses to related probes were entered into one-way repeated measures analyses of variance (ANOVAs) with three levels of prime type—congruent, incongruent, and neutral. We also examined two effects of picture-classification time. The first, the facilitation effect, was derived by subtracting mean picture-classification time on congruent trials from mean picture-classification time on neutral trials. Second, the inhibition effect was derived by subtracting mean picture-classification time for neutral trials from mean picture-classification time for incongruent trials. To evaluate the degree of facilitation versus inhibition of discourse comprehension attributable to iconic gestures, we conducted planned *t* tests comparing neutral with congruent and neutral with incongruent trials.

Finally, to explore further how the discriminability of individual items contributed to the overall pattern of results, we analyzed classification times from approximately the top and bottom 25% of items ranked according to mean ratings from the earlier norming study. Thus, strongly congruent and incongruent items were rated as such by nearly all norming participants, whereas weakly congruent and incongruent items were less consistently rated. If speech-gesture congruency played a driving role in modulating participants' behavior, we would expect rating strength to modulate the size of congruency effects—that is, we would expect larger congruency effects on classification times from strongly versus weakly rated items. This prediction was tested via a two-way repeated measures ANOVA with the factors congruency and relationship strength, as well as by follow-up *t* tests.

Span tasks and gesture-sensitivity measure

Movement span. The movement span task was administered after the picture-relatedness task. Participants stood in front of a laptop in a quiet room and watched short, soundless videos of novel movements that were grouped in successively increasing sequences ranging in length from one to five items. After the presentation of a sequence, participants were prompted to mirror what they had just seen using their own bodies. Performance was videotaped by an experimenter who was present in the room and coded off-line. A single point was awarded for each item that was correctly reproduced, irrespective of order. Half points were awarded for reenactments that clearly reflected some recollection of the target movement but deviated from it. An individual's span score was calculated as the highest consecutive level at which at least half of the total possible points were earned. If criterion was reached on a higher level after an individual's span had already been established, the final span score was increased by a half point (for additional details, see Wu & Coulson, 2014.)

Sentence span. A version of Daneman and Carpenter's (1980) sentence span task was also administered. Participants heard series of unrelated sentences and were asked to remember the sentences' final words, which they were prompted to recall at the end of each trial. Filler trials with comprehension questions were included to encourage attention to the meaning of all sentences as participants held final words in memory through internal repetition. A final sentence-span score was calculated on the basis of the highest consecutive level at which all sentence final words were accurately recalled on at least two of the three trials in a block. This task tends to prompt subvocal articulation and likely engages verbal WM.

Speech-gesture sensitivity (posttest). After completing all other tasks, participants were presented with discourse primes (excluding neutral items) a second time and asked to explicitly classify the speech and gestures in each one as congruent or incongruent. We computed *d′*—a measure of signal detection—from each participant's accuracy.

Analysis. To explore the relationship between WM abilities and speech-gesture integration, we computed Spearman's rank correlation coefficients between span scores, *d′* on the posttest, and classification-time contrasts on the picture-relatedness task (neutral versus congruent and neutral versus incongruent). Outcomes were validated with two-tailed permutation tests.

Additionally, movement-span scores were used to create KWM groups—individuals who reached criterion at Level 3 or below were classified as low, and those whose movement-span scores ranged from 3.5 to 5 were classified as high. To compare the magnitudes of facilitation and inhibition between groups, we performed a repeated measures ANOVA testing for a Group × Congruency interaction with Greenhouse-Geisser correction (original degrees of freedom are reported for clarity). Picture-classification times were subsequently analyzed separately within each group.

Finally, to examine whether KWM abilities were related to increased ease in understanding difficult-to-interpret gestures, we conducted a post hoc analysis of picture-classification times sorted by normative ratings of each discourse prime. Trials were binned according to five categories of discourse: strongly congruent (congruent speech and gestures that were consistently deemed as such), weakly congruent (congruent speech and gestures that were sometimes confused with incongruent or

deemed only weakly congruent), weakly incongruent and strongly incongruent (speech and gestures that were rated as weakly or strongly incongruent, respectively), and finally, the neutral, gesture-free counterparts to each of these trials. Again, an omnibus mixed-design 2 × 2 × 2 ANOVA was performed with these factors: group (high KWM, low KWM; between participants), congruency (congruent, incongruent; within participants), and relationship strength (strong, weak; within participants). Follow-up tests were subsequently conducted within each level of group.

Results

Picture-relatedness task

Accuracy rates. More than 90% of probes were accurately classified in all three conditions (Table 1). Nonetheless, analysis revealed a significant main effect of gesture congruence, $F(2, 166) = 8.5$, $p < .05$. Planned comparisons indicated that pictures primed by congruent discourse were classified more accurately than those preceded by neutral primes, $t(83) = 3.2$, $p < .05$. No reliable difference in classification accuracy was found between pictures preceded by neutral and incongruent discourse primes, $t(83) = -0.8$, n.s.

Picture-classification times. Average picture-classification times can be seen in Figure 4 and Table 1. Analysis revealed a main effect of gesture congruence both by subjects, $F(2, 166) = 13.4$, $p < .05$, and by items, $F(2, 276) = 8.76$, $p < .05$. Planned comparisons indicated that participants responded significantly faster to pictures preceded by congruent than by neutral primes—subjects: $t(83) = -2.08$, $p < .05$; items: $t(138) = -2.12$, $p < .05$—and significantly more slowly to pictures preceded by incongruent than by neutral primes—subjects: $t(83) = 2.9$, $p < .05$; items: $t(138) = 2$, $p < .05$.

The analysis of strongly versus weakly rated items revealed that speech-gesture congruency effects were greater for discourse primes that were rated as strongly congruent or strongly incongruent, $F(1, 76) = 5.3$, $p < .05$. Among strongly congruent and strongly incongruent items, probes following congruent primes were classified 185 ms faster on average than those following incongruent ones ($SD = 256$), $t(38) = -4.5$, $p < .05$. By contrast, among the weakly congruent and weakly incongruent items, only a 53-ms (numerical) benefit on trials in the congruent condition was observed ($SD = 324$), $t(38) = -1$, n.s.

Correlation tests

The degree to which congruent gestures facilitated picture-probe classification speed was positively correlated with movement-span scores ($\rho = 0.3$, $p < .025$; Fig. 5, Table 2). However, no relation was detected between movement-span scores and the magnitude of inhibition ($\rho = 0.09$, n.s.). Sentence-span scores did not reliably relate to either measure of picture-classification time (i.e., facilitation or inhibition), nor were sentence-span scores significantly related to movement-span scores (see Table 2).

Table 1. Mean Accuracy and Picture-Classification Time on the Picture-Relatedness Task

Prime type	Mean accuracy (% correct)	Mean classification time (ms)
Congruent	94 (0.6)	989 (37)
Neutral	92 (0.7)	1,021 (36)
Incongruent	91 (0.8)	1,080 (44)

Note: Standard errors are given in parentheses.

KWM groups

Facilitation versus inhibition. Response latencies sorted by KWM group revealed dramatically distinct patterns of facilitation and inhibition, as confirmed by a Group × Congruency interaction, $F(2, 152) = 4.0$, $p < .05$. Main effects of gesture congruence were observed in both groups—high KWM: $F(2, 76) = 7.2$, $p < .05$; low KWM: $F(2, 76) = 12.37$, $p < .05$. Follow-up t tests revealed that individuals with large movement-span scores (high KWM) tended to be facilitated by congruent gestures—neutral versus congruent: $t(37) = -3.3$, $p < .05$—but not inhibited by incongruent gestures—neutral versus

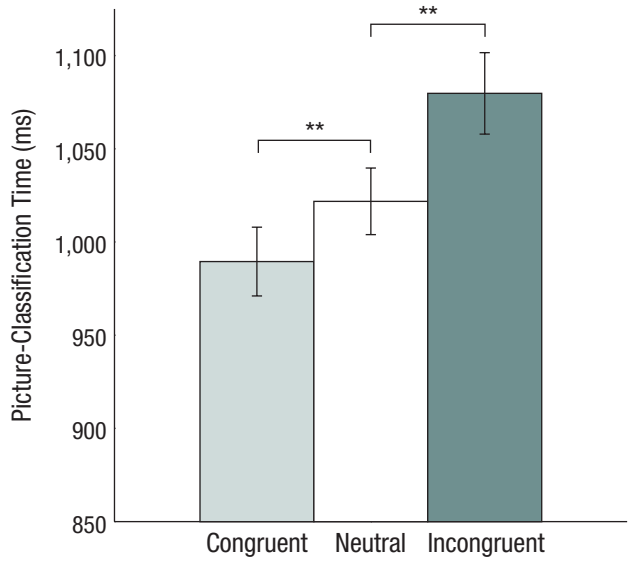


Fig. 4. Mean picture-classification time when the preceding discourse contained speech paired with congruent, incongruent, and no gestures (neutral condition). Error bars reflect 95% confidence intervals. Asterisks indicate significant differences between conditions (**$p < .01$).

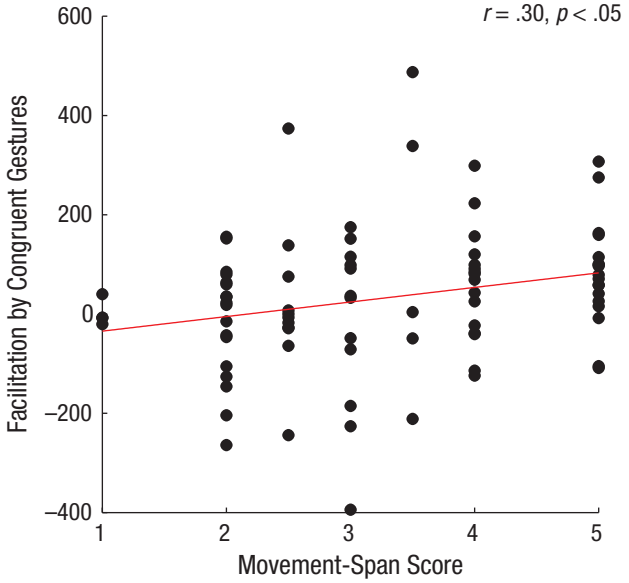


Fig. 5. Scatterplot (with best-fitting regression line) showing the correlation between the degree to which congruent gestures facilitated picture-probe classification speed and each movement-span score.

incongruent: $t(37) = 1$, n.s. By contrast, individuals with small movement-span scores (low KWM) exhibited the opposite pattern: Congruent gestures did not facilitate picture classification—neutral versus congruent: $t(37) = 0.85$, n.s.—whereas incongruent gestures led to substantially longer classification times than the neutral baseline—neutral versus incongruent: $t(37) = 4.2$, $p < .05$ (Fig. 6 and Table 3).

Levels of discriminability. A three-way interaction between KWM group, congruency, and relationship strength, $F(1, 74) = 4.5$, $p < .05$, warranted separate follow-up tests within the two levels of KWM abilities (Fig. 7). For individuals with superior KWM, the expected sensitivity to speech-gesture congruency was observed, $F(1, 37) = 14$, $p < .05$. However, these participants benefited equally from strongly and weakly congruent gestures, as evidenced by the absence of a main effect or interaction with speech-gesture relationship strength (Fs < 1, n.s.). By contrast, picture-classification times from individuals in the low-KWM group revealed a main effect of congruency, $F(1, 37) = 8$, $p < .05$, qualified by an interaction with relationship strength, $F(1, 37) = 9$, $p < .05$. Follow-up t tests revealed a marginally reliable difference in relatedness-judgment times to probes primed by congruent speech and gestures, $t(37) = -1.87$, $p = .07$, as well as a significant difference in picture-classification times on trials in which speech and gesture were incongruent, $t(37) = 2.3$, $p < .05$.

Discussion

In the experiment reported here, we found that participants' ability to reproduce meaningless actions predicted their ability to find meaning in movement. That is, participants' sensitivity to iconic gestures was systematically related to their performance on a movement span test of KWM. In keeping with the spatio-motoric hypothesis, greater KWM capacity was associated with larger facilitation by congruent speech and gestures, and it was unrelated to interference from incongruent ones (Table 2). These data suggest that listeners recruit KWM for the analysis of spatio-motoric features of the speaker's movements as they map motion segments onto conceptual knowledge about discourse referents.

Picture-classification times (Fig. 4) were consistent with substantial research indicating that iconic gestures affect language comprehension. Words are processed more effectively when preceded (Bernardis, Salillas, & Caramelli, 2008; Wu & Coulson, 2007b; Yap, So, Yap, Tan, & Teoh, 2011) or accompanied (Kelly, Creigh, & Bartolotti, 2009; Kelly, Kravitz, & Hopkins, 2004; Kelly, Ward, Creigh, & Bartolotti, 2007; Özyurek et al., 2007; Wu & Coulson, 2010b) by related than by unrelated iconic gestures. Words and pictures are primed when speech and

Table 2. Descriptive Statistics and Correlations

						Zero-order *r*			
Variable	*M*	*SD*	Minimum	Maximum	*Mdn*	1	2	3	4
1. Facilitation effect	32	142	−394	487	35	—			
2. Inhibition effect	58	182	−588	457	−41	0.50*	—		
3. Movement-span score	3.3	1.2	1.0	5.0	3.0	0.30*	0.14	—	
4. Sentence-span score	3.7	1.0	1.0	5.0	4.0	−0.05	−0.01	0.02	—

Note: The facilitation effect was derived by subtracting mean picture-classification time (in milliseconds) for congruent trials from mean picture-classification time for neutral trials. The inhibition effect was derived by subtracting mean picture-classification time for neutral trials from mean picture-classification time for incongruent trials. Scores for the movement and sentence span tasks ranged from 1 to 5.
*$p < .05$.

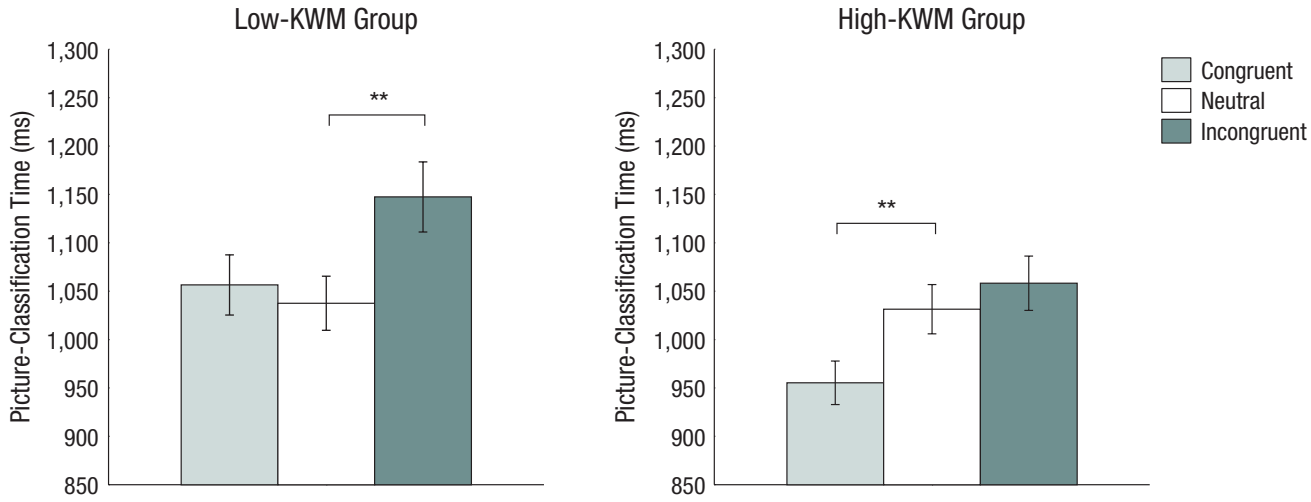


Fig. 6. Mean picture-classification time as a function of prime type. Results are shown separately for participants with low and high kinesthetic working memory (KWM). Error bars reflect 95% confidence intervals. Asterisks indicate significant differences between prime types (**$p < .01$).

gestures express consistent rather than inconsistent meanings (Holle & Gunter, 2007; Obermeier, Dolk, & Gunter, 2012; Wu & Coulson, 2010b). Relative to a gesture-free baseline, iconic gestures enhance comprehension (Beattie & Shovelton, 1999), aid word learning (Kelly, McDevitt, & Esch, 2009; Macedonia & Knösche, 2011), and guide or enhance lexical processing (Holle & Gunter, 2007; Wu & Coulson, 2010a).

The novel contribution here is the finding that taking full advantage of iconic gestures requires cognitive abilities assessed by the movement span task—presumably KWM capacity. We suggest, first, that analogous to views of verbal and spatial WM (Just & Carpenter, 1992; Shah & Miyake, 1996), KWM is a system for the maintenance and manipulation of bodily motions and, second, that listeners exploit KWM to buffer vague body movements until they can be matched and integrated with concepts evoked by the speech. The relationship shown in Figure 5 between movement-span scores and the degree of facilitation from congruent gestures presumably arises because an efficient KWM system mediates the rapid retrieval of conceptual knowledge useful for discourse comprehension.

Table 3. Mean Picture-Classification Time (ms) for Individuals With High and Low Kinesthetic Working Memory (KWM) Capacity

Prime type	Low-KWM group	High-KWM group
Congruent	1,056 (62)	955 (45)
Neutral	1,037 (72)	1,031 (51)
Incongruent	1,147 (55)	1,058 (56)

Note: Standard errors are given in parentheses.

More puzzling, however, is the question of how quantitative differences in the capacity of KWM might give rise to the disparate performance patterns shown in Figure 6. When pictures were primed by speech alone, classification times were similar across the two groups, which suggests that the observed differences in performance profiles stem from participants' ability to process multimodal discourse primes. Clearly, both KWM groups worked to integrate speech and gestures, in keeping with studies suggesting that listeners automatically engage in speech-gesture integration (Kelly, Creigh, & Bartolotti, 2009; Kelly, Özyurek, & Maris, 2010). In fact, given their longer overall response latencies on multimodal trials, participants with low KWM appear to have worked even harder than participants with high KWM to integrate speech and gestures.

Notably, Figure 7 reveals that strongly congruent speech and gestures led to facilitation in both KWM groups, whereas only participants with high KWM benefited from weakly congruent speech and gestures. Perhaps because of rapid degradation or ineffective manipulation of stored information about body movements, participants with low KWM experienced difficulty processing weakly congruent gestures. Especially in cases in which the meaning of a body movement is not readily apparent, it appears that KWM plays the important role of buffering critical information until a more robust representation can be constructed on the basis of further downstream input. Hence, it is hardly surprising that participants with low KWM often misinterpreted weakly congruent speech-gesture mappings, causing any facilitation from correctly understood trials to be washed out relative to baseline (Fig. 7).

KWM comes into play when discourse contains incongruent gestures as well. Although both groups may have

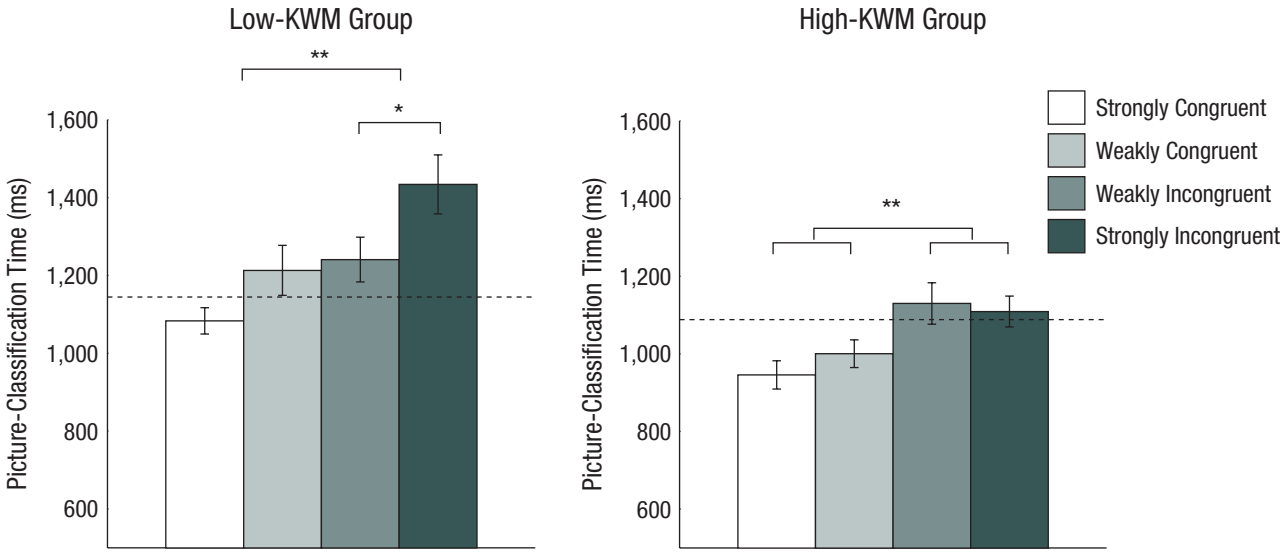


Fig. 7. Mean picture-classification time on trials for each normative rating of discourse primes. Results are shown separately for participants with low and high kinesthetic working memory (KWM). The dashed line indicates the mean decision time on trials with neutral discourse primes. Error bars reflect 95% confidence intervals. Asterisks indicate significant differences between rating types (*$p < .05$, **$p < .01$).

experienced inhibition on incongruent trials, the high-KWM group exhibited only a modest nonsignificant trend in this direction—likely because superior KWM abilities allowed these participants to rapidly determine the irrelevance of incongruent gestures and listen only to the speech. This hypothesis is consistent with the finding that performance on the movement span task correlated positively with increased sensitivity to speech-gesture congruency on the posttest ($\rho = 0.35$, $p < .05$; see also Wu & Coulson, 2014). That is, individuals who most consistently discriminated between congruent and incongruent gestures also tended to score the highest on the movement span task.

These data suggest that the low-KWM group had a limited ability to differentiate congruent from incongruent speech and gestures and, consequently, formulated off-target interpretations of incongruent gestures rather than ignoring them. This tendency was particularly detrimental on the strongly incongruent trials, on which gestures may have led these participants to construct situation models that differed substantially from the picture probes—although they eventually judged correctly that the probes were related to the discourse primes.

On the other hand, attempting to interpret the weakly incongruent gestures was less problematic, and, accordingly, participants with lower KWM experienced less inhibition on these trials. Indeed, the weakly incongruent gestures may have been judged as such because they communicated concepts that could be forced to cohere with the concurrent speech. For example, in one weakly incongruent trial, the utterance "the handle at the top of the stove that can be pulled out" was paired with an incongruent gesture in which the speaker indicated the spatial extent of an object similar in size and shape to a small stove (whereas the original congruent gesture depicted the handle).

Notably, our finding that participants with low KWM exhibited interference supports the claim that poor performance on span tasks—that is, those requiring participants to maintain recently encoded information in the face of competition from other aspects of the task—is at least partially attributable to a domain-general susceptibility to interference (Engle & Kane, 2004). Indeed, KWM abilities measured here may coincide with individual differences along related, but nevertheless distinct, cognitive dimensions—such as greater perceptual sensitivity or attention to body movements or the ability to rapidly shift attention. This view fits with noncomponential models of WM emphasizing the role of executive attentional control in determining WM capacity and tapping processes that correlate with measures of general intelligence (Cowan, 2008; Engle, 2002). Rather than being viewed as a collection of independent, dedicated subcomponents, WM has been framed simply as hierarchically embedded subsets of information in long-term memory that become activated under the focus of attention (Cowan, 1999). However, in the present study, the ability to remember and rehearse verbal information did not relate either to movement span or to our measure of multimodal discourse comprehension, which suggests that movement span indexes abilities at least partially dissociable from sentence span—in keeping with our claim that it taps KWM.

Two important conclusions emerge from the analysis of KWM groups. First, although listeners likely engage in speech-gesture integration automatically, attention to

gestures deemed irrelevant can be inhibited—particularly in individuals with superior KWM abilities. Second, that congruent gestures enhanced message-level comprehension in participants with high KWM reinforces the view that KWM mediates the process of mapping meaningful body movements to concepts.

Conclusion

Cognitive systems important for remembering and reproducing meaningless body movements also contribute to the processes listeners use in understanding multimodal discourse. The present study revealed a relationship between kinesthetic, but not verbal, WM capacity in the interpretation of gestures. Because gesture meaning is not always immediately apparent, KWM may allow the listener to maintain gestural information until it can be matched to concepts evoked by the speech. Speech-gesture integration processes are not monolithic, however—as iconic gestures influence individuals in different ways. The degree of facilitation by congruent gestures was systematically related to movement span, as individuals with higher movement-span scores tended to benefit more from congruent gestures—even weakly congruent gestures whose relationship to the discourse was difficult to discern. Moreover, the low-KWM group tended to experience interference from all incongruent gestures, and even from weakly congruent ones, as if such gestures prompted the retrieval of inappropriate information about discourse referents.

Author Contributions

Y. C. Wu developed the study concept. Both authors designed and implemented the experiment, and both collaborated on writing the manuscript and approved the final manuscript for submission.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was supported by National Science Foundation Grant No. BCS-0843946 to S. Coulson.

Supplemental Material

Additional supporting information can be found at http://pss.sagepub.com/content/by/supplemental-data

References

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York, NY: Academic Press.

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

Bernardis, P., Salillas, E., & Caramelli, N. (2008). Behavioral and neurophysiological evidence of semantic interaction between iconic gestures and words. *Cognitive Neuropsychology*, *25*, 1114–1128.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge, England: Cambridge University Press.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, *169*, 323–338.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo–spatial memory. *Neuropsychologia*, *37*, 1189–1199.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19–23.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 145–199). New York, NY: Academic Press.

Goodwin, C. (2003). The body in action. In J. Coupland & R. Gwyn (Eds.), *Discourse, the body, and identity* (pp. 19–42). New York, NY: Palgrave Macmillan.

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

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, *19*, 1175–1192.

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.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

Kelly, S. D., Creigh, P., & Bartolotti, J. (2009). Integrating speech and iconic gestures in a Stroop-like task: Evidence for automatic processing. *Journal of Cognitive Neuroscience*, *22*, 683–694.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain & Language*, *89*, 253–260.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*, 313–334.

Kelly, S. D., Özyurek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, *21*, 260–267.

Kelly, S. D., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and

speech during comprehension. *Brain & Language*, *101*, 222–233.

Klauer, K. C., & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of Experimental Psychology: General*, *133*, 355–381.

Macedonia, M., & Knösche, T. R. (2011). Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education*, *5*, 196–211.

Obermeier, C., Dolk, T., & Gunter, T. (2012). The benefit of gestures during communication: Evidence from hearing and hearing-impaired individuals. *Cortex*, *48*, 857–870.

Özyurek, A., Willems, R. M., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*, 605–616.

Posner, M. (1967). Characteristics of visual and kinesthetic memory codes. *Journal of Experimental Psychology: General*, *75*, 103–107.

Seemüller, A., Fiehler, K., & Rösler, F. (2011). Unimodal and crossmodal working memory representations of visual and kinesthetic movement trajectories. *Acta Psychologica*, *136*, 52–59.

Seemüller, A., Müller, E. M., & Rösler, F. (2012). EEG-power and -coherence changes in a unimodal and crossmodal working memory task with visual and kinesthetic stimuli. *International Journal of Psychophysiology*, *83*, 87–95.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*, 4–27.

Smyth, M., Pearson, N. A., & Pendleton, L. R. (1988). Movement and working memory: Patterns and positions in space. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *40*, 497–514.

Smyth, M., & Pendleton, L. (1989). Working memory for movements. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *41*, 235–250.

Smyth, M., & Pendleton, L. (1990). Space and movement in working memory. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *42*, 291–304.

Streeck, J. (2002). A body and its gestures. *Gesture*, *2*, 19–44.

Streeck, J. (2008). Depicting by gesture. *Gesture*, *8*, 285–301.

Wu, Y. C., & Coulson, S. (2007a). How iconic gestures enhance communication: An ERP study. *Brain & Language*, *101*, 234–245.

Wu, Y. C., & Coulson, S. (2007b). Iconic gestures prime related concepts: An ERP study. *Psychonomic Bulletin & Review*, *14*, 57–63.

Wu, Y. C., & Coulson, S. (2010a). Gestures modulate speech processing early in utterances. *NeuroReport*, *21*, 522–526.

Wu, Y. C., & Coulson, S. (2010b, November). *Iconic gestures facilitate word and message processing: The multi-level integration model of audio-visual discourse comprehension*. Paper presented at the Neurobiology of Language Conference, San Diego, CA. Abstract retrieved from http://www.neurolang.org/programs/ScientificProgramNLC2010.pdf

Wu, Y. C., & Coulson, S. (2014). A psychometric measure of working memory for configured body movement. *PLoS ONE*, *9*(1), Article e84834. Retrieved from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0084834

Yap, D.-F., So, W.-C., Yap, J.-M. M., Tan, Y.-Q., & Teoh, R.-L. S. (2011). Iconic gestures prime words. *Cognitive Science*, *35*, 171–183.