Psychological Science

http://pss.sagepub.com/

Two Sides of the Same Coin: Speech and Gesture Mutually Interact to Enhance Comprehension

Spencer D. Kelly, Asli Özyürek and Eric Maris

Psychological Science 2010 21: 260 originally published online 22 December 2009

DOI: 10.1177/0956797609357327

Research Article

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
21(2) 260–267

DOI: 10.1177/0956797609357327
http://pss.sagepub.com
SAGE

Two Sides of the Same Coin: Speech and Gesture Mutually Interact to Enhance Comprehension

Spencer D. Kelly[1], **Aslı Özyürek**[2,3], **and Eric Maris**[4,5]
[1]Department of Psychology, Neuroscience Program, Colgate University; [2]Department of Linguistics, Radboud University Nijmegen; [3]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; [4]Donders Institute for Brain, Cognition and Behaviour, Center for Cognition, Radboud University Nijmegen; and [5]Donders Center for Cognitive Neuroimaging, Radboud University Nijmegen

Abstract

Gesture and speech are assumed to form an integrated system during language production. Based on this view, we propose the integrated-systems hypothesis, which explains two ways in which gesture and speech are integrated—through *mutual* and *obligatory* interactions—in language comprehension. Experiment 1 presented participants with action primes (e.g., someone chopping vegetables) and bimodal speech and gesture targets. Participants related primes to targets more quickly and accurately when they contained congruent information (speech: "chop"; gesture: chop) than when they contained incongruent information (speech: "chop"; gesture: twist). Moreover, the strength of the incongruence affected processing, with fewer errors for weak incongruities (speech: "chop"; gesture: cut) than for strong incongruities (speech: "chop"; gesture: twist). Crucial for the integrated-systems hypothesis, this influence was bidirectional. Experiment 2 demonstrated that gesture's influence on speech was obligatory. The results confirm the integrated-systems hypothesis and demonstrate that gesture and speech form an integrated system in language comprehension.

Keywords

speech, language, communication, iconic gestures, action, comprehension, semantic processing, multimodal, mutual, obligatory, integrated-systems hypothesis

Received 2/20/09; Revision accepted 6/2/09

It is no longer news to point out that people gesture when they speak. Furthermore, the idea that gesture and speech form an integrated system sounds more plausible than ever (Bernardis & Gentilucci, 2006; Clark, 1996; Kendon, 2004; Kita & Özyürek, 2003; McNeill, 1992, 2005, in press), especially in the light of recent discoveries concerning the links between language and action (Fischer & Zwaan, 2008; Masson, Bub, & Warren, 2008; Pulvermüller, 2005; Willems & Hagoort, 2007). Although theories on gesture-speech integration originated in the realm of language production (e.g., Bernardis & Gentilucci, 2006; Kita & Özyürek, 2003; McNeill, 1992, 2005), it has become standard practice to assume that this integrated relationship extends also to the comprehension domain (Bernardis, Salillas, & Caramelli, 2008; Holle & Gunter, 2007; Kelly, Kravitz, & Hopkins, 2004; Kelly, Ward, Creigh, & Bartolotti, 2007; Özyürek, Willems, Kita, & Hagoort, 2007; Wu & Coulson, 2007). However, to date, research on gesture comprehension has yet to clearly articulate how it is rooted in the specific theoretical claims made in the arena of production. In the present article, we attempt to bridge this gap by introducing a new hypothesis—the integrated-systems hypothesis—that explicitly links theories on gesture production to empirical findings on gesture comprehension.

We focus on two theoretical claims about gesture-speech production that are particularly relevant to processes involved in language comprehension: the bidirectional influence of gesture and speech and the obligatory integration of the two modalities. For example, Kita and Özyürek (2003) argued that speech affects what people produce in gesture and that gesture, in turn, affects what people produce in speech. In other words, the two modalities bidirectionally interact during language production. Moreover, McNeill (in press) argued that this interaction was so fundamental that under many circumstances, gesture and speech were obligatorily coupled. That is, producing a gesture often requires speech. These two claims

Corresponding Author:
Spencer D. Kelly, Department of Psychology, Neuroscience Program, 13 Oak Dr., Colgate University, Hamilton, NY 13346
E-mail: skelly@colgate.edu

are central to the general theory that gesture and speech are actually part and parcel of language—that is, they together *constitute* language (McNeill, 1992, 2005).

This research in language production is useful for guiding hypotheses about language comprehension. Indeed, although previous research has demonstrated that gesture and speech combine during comprehension (for a recent review, see Kelly, Manning, & Rodak, 2008), the precise nature of this interaction is unclear. Building on the above two claims in production, we introduce the integrated-systems hypothesis, positing that gesture and speech mutually and obligatorily interact with one another to enhance language comprehension; that is, gesture influences the processing of speech, speech influences the processing of gesture, and this integration is mandatory.

As an introduction to the basic design of our experiments, consider the following real-life example: In a much-publicized videoclip—number two on David Letterman's Top 10 George W. Bush moments—President Bush said the following in a speech, "The left hand now knows what the right hand is doing." This by itself is a fine idiom, but what the president simultaneously did with his hands was inexplicable: He gestured first with his right hand and then with his left hand. This multimodal "Bushism" raises some simple but interesting questions. Does a message with incongruent speech and gesture disrupt comprehension for listeners? Would a message with congruent speech and gesture enhance understanding? The present study provides affirmative answers to these questions and proposes the integrated-systems hypothesis as a mechanism.

To test the integrated-systems hypothesis, we conducted two experiments using a simple priming paradigm in which videos presented action primes (e.g., chopping vegetables) followed by targets comprising speech and iconic gestures.[1] In each video, at least one modality of the targets was related to the prime, but the congruence between the gesture and speech varied within the targets (as in the Bush example). In some cases, the gesture and speech were congruent (speech: "chop"; gesture: chop), and in other cases, they were incongruent. Unique to the present study is the way in which they were incongruent: In one condition, two modalities were weakly incongruent (speech: "chop"; gesture: cut), and in the other, they were strongly incongruent (speech: "chop"; gesture: twist). This novel incongruence paradigm allowed us to investigate how different levels of semantic incongruence modulated gesture-speech integration. The task was to identify whether any information in the target (speech or gesture) was related to the prime (Experiment 1), or whether information contained only in the spoken portion of the target was related to the prime (Experiment 2).

Using this basic paradigm, we made two predictions from the integrated-systems hypothesis. (a) Gesture and speech should mutually interact, with incongruent gestures disrupting the ability to relate speech targets to primes and incongruent speech disrupting the ability to relate gesture targets to primes (Experiment 1). (b) This integration should be obligatory, so that even when the task does not require any attention to one modality (e.g., gesture), it should nevertheless affect how people relate the other modality (e.g., speech) to the prime (Experiment 2).

Experiment 1

Method

Participants. Twenty-nine right-handed college undergraduates (13 males and 16 females) participated in Experiment 1. All were native English speakers and were recruited from an introductory psychology course.

Materials. The stimuli in the experiment were 1-s videos of a common action prime (e.g., chopping vegetables, washing dishes, or hammering a nail), followed by a 500-ms black screen, and finished by a verbal or gestural target (for 1 s). A male actor produced the prime, and a female actor produced the speech and gesture targets. To create the targets, the woman was asked to watch the action videos and spontaneously and naturally produce a word and gesture that matched the video.

For Experiment 1, just over half of the videos conveyed information (in speech or gesture) that was related to the prime, and just under half conveyed unrelated information (this second group of stimuli served as filler items and was not part of the analyses). In the baseline condition, the gesture (e.g., gesturing chop) and speech (e.g., saying "chop") conveyed congruent information and were both related to the prime (e.g., chopping vegetables). The four other conditions were created using a new levels-of-incongruence paradigm, in which only one piece of information (speech or gesture) was related to the prime, but the level of semantic incongruence between gesture and speech varied within the targets. For example, in the two speech-target conditions, the verbal portion of the target was related to the prime, but the gesture was either strongly incongruent (e.g., gesturing twist) or weakly incongruent (e.g., gesturing cut). In the two gesture-target conditions, the gestural portion of the target was related to the prime, but the speech was either strongly incongruent (e.g., saying "twist") or weakly incongruent (e.g., saying "cut"). Figure 1 provides an example of the action prime and five targets for the experimental conditions.

Note that this incongruence paradigm is novel because, unlike previous research comparing congruent and incongruent gesture and speech pairs, it allows for an additional comparison within incongruent conditions. In this way, we obtain a stronger experimental control than in the comparison between congruent and incongruent gesture and speech pairs, which differ both in their relation to the prime and in their relation to one another. In contrast, weakly and strongly incongruent speech and gesture pairs differ only in their semantic relation to one another.

For the experimental trials, there were 16 videos for each of the five conditions, yielding 80 experimental items. In

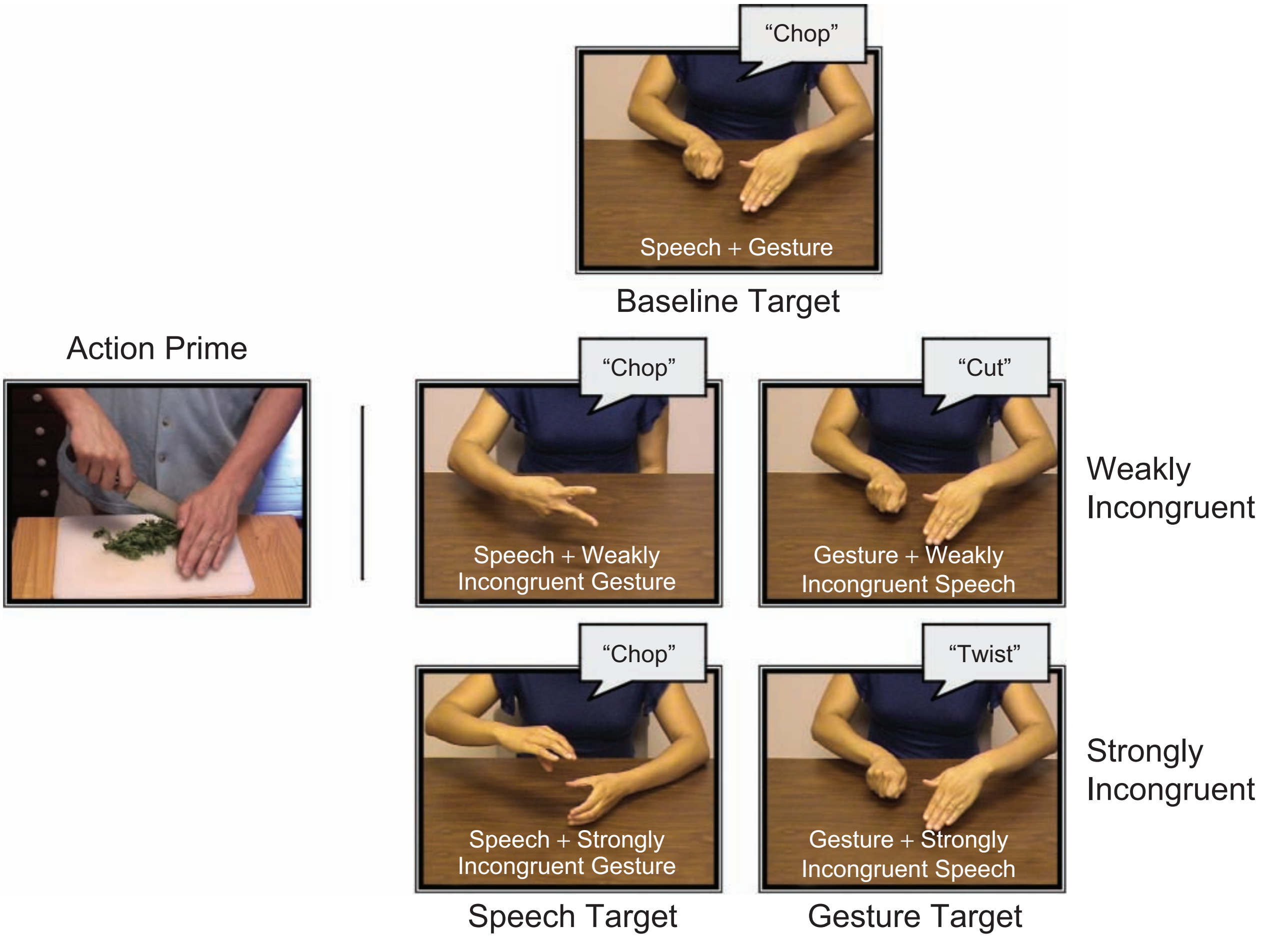

Fig. 1. Illustration of the stimuli for both experiments. The stimuli consisted of a video of an action prime (shown for 1 s), followed by a 500-ms black screen, and finished by a 1-s video of a target. For the baseline target, the gesture and speech conveyed congruent information, and both were related to the prime. In the two speech-target conditions, the verbal portion of the target was related to the prime, but the gesture was either weakly incongruent (e.g., gesturing cut while saying "chop") or strongly incongruent (e.g., gesturing twist while saying "chop"). In the two gesture-target conditions, the gestural portion of the target was related to the prime, but the speech was either weakly incongruent (e.g., saying "cut" while gesturing chop) or strongly incongruent (e.g., saying "twist" while gesturing chop).

addition, there were 64 fillers items (no target information was related to the prime), yielding a total of 144 videos.

Procedure. On arrival at the study location, participants were directed to an individual testing room with a computer and keyboard. Participants were instructed to press one keyboard button ("no" responses) if no part of the target—speech or gesture—was related to the prime and to press a different button ("yes" responses) if any part of the target (speech or gesture) was related to the prime. In this way, participants would press "yes" for all of the experimental stimuli and "no" for all of the filler stimuli. The instructions emphasized the importance of both speed and accuracy in responding to the trials. Before participants did the actual experiment, they were given six practice trials and asked if they had any questions. The entire procedure lasted approximately 35 min.

Response times (in milliseconds) and error rates (proportion incorrect) were analyzed to determine how well participants could relate the speech and gesture targets to the action primes. For the response time measure, outliers two standard deviations from the mean were removed (Ratcliff, 1993), and only responses that were correct were entered into the final analyses.

Prediction and analyses. We predicted that gesture and speech would bidirectionally interact when relating targets to primes. We tested this prediction in two ways. First, as a coarse measure of bidirectional interaction, we expected that relating targets to primes would produce faster response times and fewer errors in the baseline condition (where gesture and speech conveyed congruent information) compared with the four incongruent conditions. To test this, we ran two F tests (response times and errors) comparing the baseline condition with the average of the four incongruent conditions. Second, as a more direct measure, we focused on just the four incongruent conditions and conducted a 2 (strength of incongruence: weak vs. strong) × 2 (target modality: speech vs. gesture) repeated measures analysis of variance (ANOVA) on the response times and error rates, expecting a main effect of incongruence—with strongly incongruent stimuli producing slower response times and more errors than weakly incongruent stimuli—and, crucially, no significant interaction between incongruence and target modality. This pattern would indicate that strongly incongruent gesture disrupts speech targets to a comparable degree as strongly incongruent speech disrupts gesture targets. All F tests were adjusted using the Greenhouse-Geisser correction.

Results

Analysis 1. In keeping with our prediction, the difference for the response times was significant by subjects, $F(1, 28) = 37.74$, $p < .001$, $d = 0.574$, and items, $F(1, 15) = 17.85$, $p < .001$, $d = 0.543$, with participants producing faster response times in the baseline condition than in the four incongruent conditions. Also consistent was a significant effect for the error rates, by subjects, $F(1, 28) = 70.91$, $p < .001$, $d = 0.715$, and items, $F(1, 15) = 198.82$, $p < .001$, $d = 0.930$, with participants producing fewer errors in the baseline condition than in the four incongruent conditions. Note that there was no speed-accuracy trade-off: Participants were faster and more accurate in the baseline condition (see Fig. 2).

Analysis 2. For the response times, there was not a significant main effect of incongruence by subjects, $F(1, 28) = 2.05$, n.s., or items, $F(1, 15) = 1.03$, n.s., and there was no significant interaction between target modality and incongruence by subjects, $F(1, 28) = 0.17$, n.s., or items, $F(1, 15) = 0.01$, n.s. Interestingly, although not part of our predictions, there was a significant main effect of target modality by subjects, $F(1, 28) = 5.57$, $p = .025$, $d = 0.166$, and items, $F(1, 15) = 4.52$, $p = .05$, $d = 0.232$, with speech targets producing faster response times than gesture targets.

Although the response time results did not support our prediction, the findings from the error rates did. Not only was there a significant main effect of incongruence by subjects, $F(1, 28) = 5.04$, $p = .033$, $d = 0.153$, and items, $F(1, 15) = 14.30$, $p = .002$, $d = 0.488$, with strongly incongruent pairs producing more errors than weakly incongruent pairs, but, crucially, there was not a significant interaction between target modality and congruence by subjects, $F(1, 28) = 0.03$, n.s., or items, $F(1, 15) = 0.16$, n.s. Finally, unlike the response time measure, there was not a significant main effect of target modality by subjects, $F(1, 28) = 3.34$, n.s., or items, $F(1, 15) = 0.77$, n.s. (see Fig. 2).

Discussion

The results from Experiment 1 support our first prediction. For Analysis 1, not only were participants faster to correctly identify targets when gesture and speech conveyed congruent information but they produced fewer errors as well.[2] And for Analysis 2, the error rates yielded a main effect of incongruence but no interaction between incongruence and target modality, suggesting that gestures influenced speech comprehension, speech influenced gesture comprehension, and this influence was comparable.

In addition, although it was not part of our prediction, we found that speech targets were processed faster than gesture targets. We do not want to make too much of this unexpected finding, but it is interesting that although the gestures were physically more similar to the action primes, it was speech—which was related to the primes only though linguistic convention—that was processed faster. Future research should investigate this intriguing finding further to compare the relative costs and benefits of representing actions in an imagistic (gesture) versus arbitrary (speech) fashion.

In our next experiment, we tested the second component of the integrated-systems hypothesis: whether there is an obligatory interaction between the gesture and speech at comprehension. To do this, we presented the same stimuli from Experiment 1, but the instructions were to attend only to the verbal portion of the videos. If gestures continued to influence speech processing even when they were not relevant to the task, it would be strong evidence for the second aspect of the integrated-systems hypothesis.[3]

Experiment 2

Method

Participants. A separate group of 41 right-handed college undergraduates (11 males and 30 females) participated in Experiment 2. Again, all were native English speakers and were recruited from an introductory psychology course.

Materials. The videos were the same as in Experiment 1, with one major difference. Within the experimental trials, the gesture-target conditions became filler items because the task was now focused on speech. So the new experimental trials consisted of the baseline condition and the two speech-target conditions. There were 16 videos for each of these three conditions (for a total of 48 trials). In addition, there were 64 filler items (we also dropped 32 fillers from Experiment 1 for a better balance), yielding a total of 112 videos.

Procedure. The basic procedure was the same as Experiment 1, but the instructions were different. This time, participants were told that they would be watching action primes that were followed by information conveyed in speech and gesture, but that the task concerned whether *only* the speech content was the same or different from the primes.[4] The entire procedure lasted approximately 20 min.

Prediction and analyses. Using one-way (baseline vs. weakly incongruent vs. strongly incongruent) repeated measures

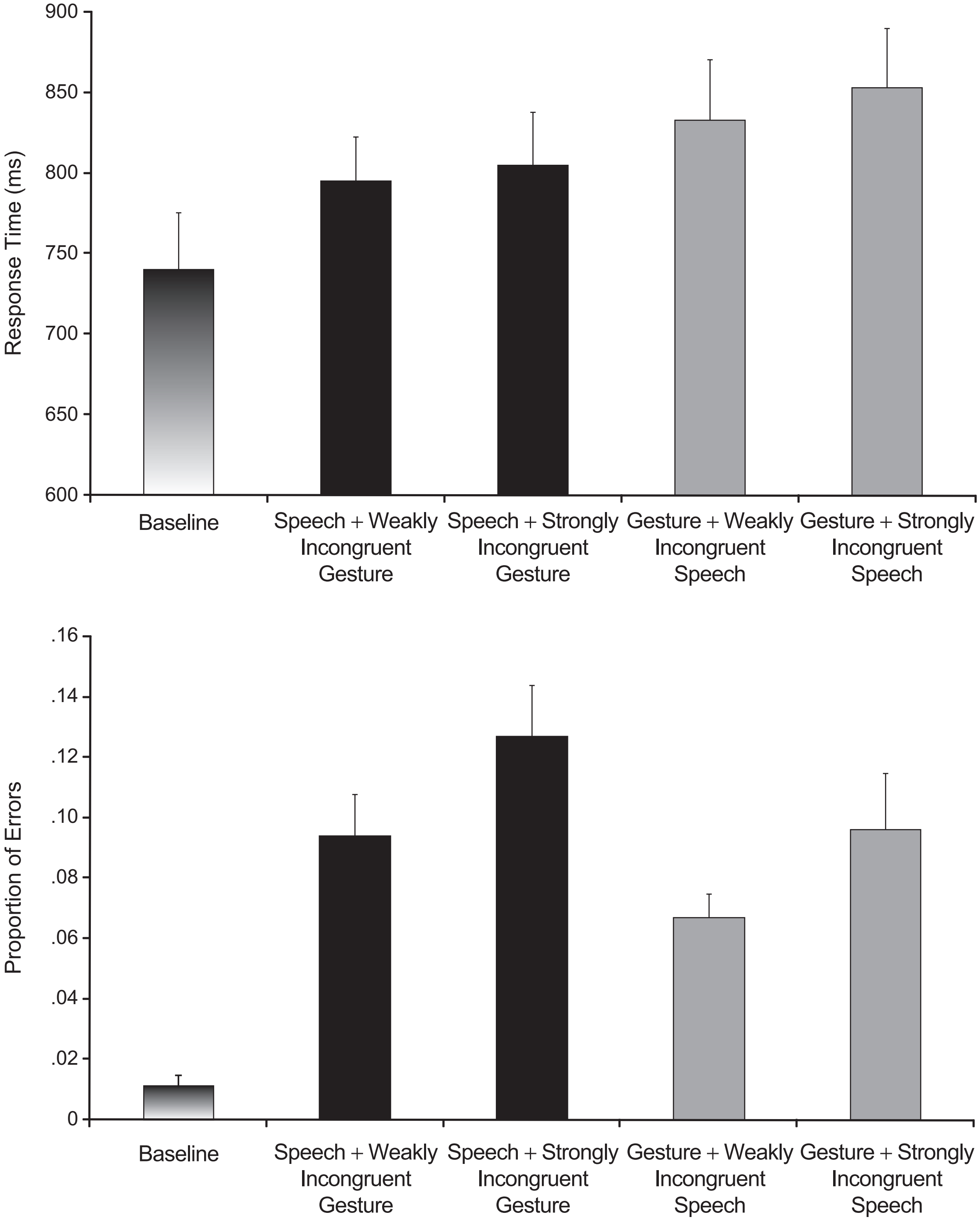

Fig. 2. Results of Experiment 1: mean response time (top panel) and proportion of errors (bottom panel) as a function of condition. Error bars represent standard errors.

ANOVAs, we predicted that if gesture and speech were obligatorily integrated, participants would be slower and/or less accurate to relate the speech targets to action primes as gesture and speech become increasingly incongruent.

Results

The ANOVA was not significant for response times by subjects, *F*(2, 80) = 1.19, n.s., or items, *F*(2, 30) = 2.02, n.s. However, the

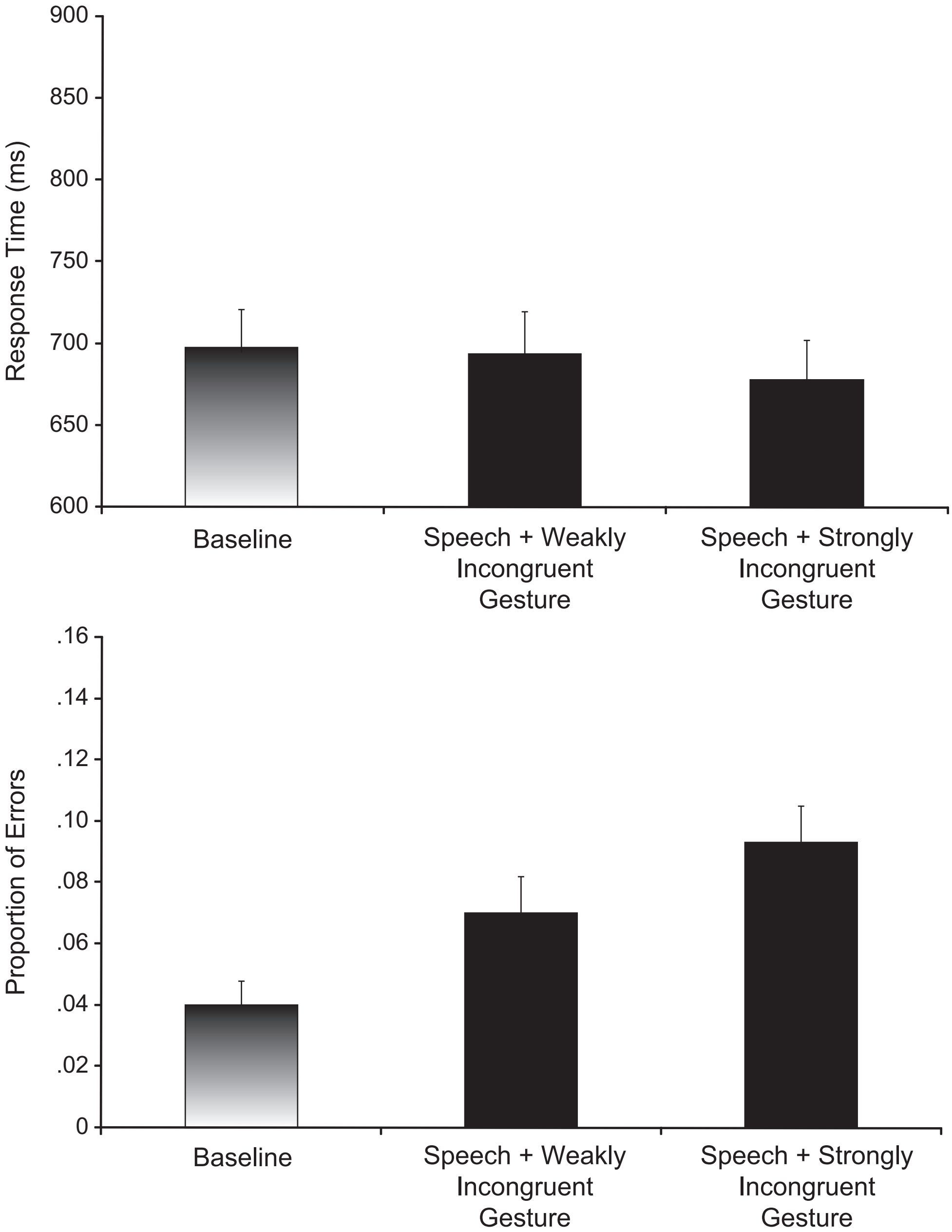

Fig. 3. Results of Experiment 2: mean response time (top panel) and proportion of errors (bottom panel) as a function of condition. Error bars represent standard errors.

there was a significant effect for error rates by subjects, $F(2, 80) = 6.70$, $p = .002$, $d = 0.143$, and items, $F(2, 30) = 20.83$, $p < .001$, $d = 0.582$. A linear trend analysis demonstrated that error rates went up as the semantic distance between speech and gesture increased, by subjects, $F(1, 40) = 13.76$, $p < .001$, $d = 0.256$, and items, $F(1, 15) = 61.03$, $p < .001$, $d = 0.803$ (see Fig. 3).

Discussion

Even when the task did not include instructions to respond to gestural information contained in the videos, accuracy scores demonstrated sensitivity to gesture by becoming decreasingly accurate as semantic incongruity with speech increased. This finding supports our second prediction.

General Discussion

The results from our two experiments have confirmed our two predictions. First, when gesture and speech convey the same information, they are easier to understand—they are faster and produce fewer errors—than when they convey different information, and this effect appears to be driven by mutual interactions, with strong incongruities between speech and gesture bidirectionally affecting integration to a greater extent than weak incongruities. Second, this integration is obligatory: People cannot help but consider one modality (gesture) when processing the other (speech).

These findings are novel because in addition to highlighting mechanisms of gesture-speech integration in comprehension, they bolster theories that the two modalities comprise an integrated system in production (Bernardis & Gentilucci, 2006; Clark, 1996; Kendon, 2004, Kita & Özyürek, 2003; McNeill, 1992, 2005, in press). Researchers have theorized that gesture and speech are so tightly integrated in production that they together constitute language (McNeill, 1992), through mutual interactions (Kita & Özyürek, 2003) and obligatory couplings (McNeill, in press). The results from the present study lend support to these claims by showing that this integrated relationship may hold for language comprehension as well.

It is also informative to connect the integrated-systems hypothesis to research in other areas demonstrating bidirectional and obligatory interactions in multimodal communication (Calvert, 2001; de Gelder & Vroomen, 2000; van Wassenhove, Grant, & Poeppel, 2005). For example, in a very relevant study on the comprehension of emotional information, de Gelder and Vroomen (2000) demonstrated that affective information conveyed through the face and voice is integrated in a bidirectional fashion and that this integration is mandatory. They concluded that this tight relationship may reflect the basic multimodal architecture of the human brain, which may be designed to optimally process and integrate information from across modalities (see also Calvert, Spence, & Stein, 2004). These studies, coupled with recent work showing that gesture and speech share similar neural mechanisms during language comprehension (Willems, Özyürek, & Hagoort, 2007), give credence to the claim that the two modalities form a fundamentally integrated system in both language production and comprehension.

Before we conclude, it is worth commenting on a few other aspects and implications of the results. Although previous research in the area of cognitive neuroscience has demonstrated that people semantically integrate gesture and speech differently when they are congruent versus incongruent (Bernardis et al., 2008; Cornejo et al., 2009; Kelly et al., 2004, 2007; Özyürek et al., 2007), no study has systematically explored how different degrees of semantic incongruity affect processing. Using our unique levels-of-incongruence paradigm, we were able to show that when gesture and speech convey incongruent information, the strength of the semantic relationship matters. This finding suggests that gesture and speech may be semantically integrated in a graded fashion, much in the same way that the degree of semantic expectancy of a word modulates how easily it is integrated into a sentence (Kutas & Hillyard, 1984).

Related to this point, note that in Experiment 1, incongruent speech was not more disruptive than incongruent gesture (i.e., we found a main effect for the error rates of incongruence but no interaction with target modality). This is a provocative finding. Almost all previous research on gesture comprehension has treated gesture as a context for speech (for a review, see Kelly et al., 2008), but we have shown here that speech is also a context for gesture, suggesting that the two modalities may co-determine meaning during language comprehension (see also Kelly, Barr, Church, & Lynch, 1999). This fits nicely with the view that gesture and speech combine to form a composite signal in communication (Clark, 1996) or, to put it another way, that they are simply two sides of the same coin: language.

It is worth reiterating that incongruent speech and gesture disrupted understanding to a very great extent—indeed, compared with performance during the congruent condition, participants during the incongruent conditions in Experiment 1 made on average 8 times as many errors when gesture and speech were weakly incongruent and 11 times as many errors when they were strongly incongruent. This finding has obvious implications for the multimodal conditions that are most conducive to effectively communicate information in everyday face-to-face interactions. But it is also interesting in the light of research demonstrating that certain incongruities between gesture and speech (so called mismatches) actually facilitate learning and development (Goldin-Meadow, 2003; Singer & Goldin-Meadow, 2005). The present results raise the intriguing possibility that gesture may have different functions on different time scales: For quickly and accurately understanding messages in the moment, congruencies between gesture and speech are by far the most effective; but for shaking up knowledge states and prodding learners over time, a certain degree of incongruence between gesture and speech (mismatches) may be optimal.

To conclude, in accordance with the predictions of the integrated-systems hypothesis, gesture and speech interact in a mutual and obligatory fashion, and when conveying the same message, they greatly enhance understanding. This not only helps to elucidate mechanisms of how gesture and speech are semantically integrated during language comprehension but also has implications for everyday life—if you really want to make your point clear and readily understood, let your words *and* hands do the talking.

Acknowledgments

We would like to thank the anonymous reviewers for their helpful suggestions on previous versions of this manuscript. We are also grateful to the feedback we received at the Nijmegen Gesture Center (particularly from Pamela Perniss) while the first author was on sabbatical at the Max Planck Institute for Psycholinguistics (The Netherlands). Finally, we thank Ron Crans for technical assistance in software creation and the students in Colgate's Research Methods course for assistance with data collection.

Declaration of Conflicting Interests

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

Notes

1. Iconic gestures convey information about objects, actions, and events that semantically overlap with, and often complement, the spoken utterance (McNeill, 1992).

2. At this point, many readers might wonder how congruent gesture and speech compare to gesture or speech alone. Although we do not have space to report the results in full, we conducted an experiment on a separate group of 69 participants and found that although there was a floor effect for error rates, congruent gesture-speech targets produced significantly faster response times than were found in a speech-only condition, Dunn-Sidak $t(3, 68) = 4.26$, $p < .001$, or a gesture-only condition, Dunn-Sidak $t(3, 68) = 13.58$, $p < .001$. Interestingly, the speech condition produced faster response times than did the gesture condition, Dunn-Sidak $t(3, 68) = 5.90$, $p < .001$. In this way, it is safe to say that congruent gesture-speech pairs enhance understanding, compared with unimodal and incongruent bimodal messages.

3. Note that we investigated only whether gesture is obligatorily processed when attending to speech, and not the other way around. We made this decision because it would allow us to more directly relate our results to previous research, which has almost exclusively focused on gesture as a context for speech.

4. We also collected data from a separate set of 56 participants whom we explicitly told to ignore hand gestures. (The present manipulation did not go that far—it merely defined the task as one focused on speech.) Although not reported here, the results from that experiment exactly mirror what we found in the present experiment.

References

Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia*, *44*, 178–190.

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

Calvert, G.A. (2001). Cross-modal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, *11*, 1110–1123.

Calvert, G.A., Spence, C., & Stein, B.E. (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT Press.

Clark, H.H. (1996). *Using language*. Cambridge, England: Cambridge University Press.

Cornejo, C., Simonetti, F., Ibáñez, A., Aldunate, N., Ceric, F., Lopez, V., et al. (2009). Gesture and metaphor comprehension: Electrophysiological evidence of cross-modal coordination by audiovisual stimulation. *Brain and Cognition*, *70*, 42–52.

de Gelder, B., & Vroomen, J. (2000). Perceiving emotions by ear and by eye. *Cognition & Emotion*, *14*, 289–311.

Fischer, M.H., & Zwaan, R.A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, *61*, 825–850.

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

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

Kelly, S.D., Barr, D.J., Church, R.B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, *40*, 577–592.

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

Kelly, S.D., Manning, S., & Rodak, S. (2008). Gesture gives a hand to language and learning: Perspectives from cognitive neuroscience, developmental psychology and education. *Language and Linguistics Compass*, *2*, 1–20.

Kelly, S.D., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language*, *101*, 222–233.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge, England: Cambridge University Press.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*, 16–32.

Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163.

Masson, M.E.J., Bub, D.N., & Warren, C.M. (2008). Kicking calculators: Contribution of embodied representations to sentence comprehension. *Journal of Memory and Language*, *59*, 256–265.

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

McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

McNeill, D. (in press). Gesture. In P.C. Hogan (Ed.), *Cambridge encyclopedia of the language sciences*. Cambridge, England: Cambridge University Press.

Özyürek, 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.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*, 576–582.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*, 510–532.

Singer, M.A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science*, *16*, 85–89.

van Wassenhove, V., Grant, K.W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences, USA*, *102*, 1181–1186.

Willems, R.M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language*, *101*, 278–289.

Willems, R.M., Özyürek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex*, *17*, 2322–2333.

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