Available online at www.sciencedirect.com
ScienceDirect
Brain and Language 101 (2007) 222–233
Brain and Language
www.elsevier.com/locate/b&l

An intentional stance modulates the integration of gesture and speech during comprehension ☆

Spencer D. Kelly *, Sarah Ward, Peter Creigh, James Bartolotti

Neuroscience Program, Department of Psychology, 13 Oak Dr, Colgate University, Hamitlon, NY 13346, USA

Accepted 25 July 2006
Available online 25 September 2006

Abstract

The present study investigates whether knowledge about the intentional relationship between gesture and speech influences controlled processes when integrating the two modalities at comprehension. Thirty-five adults watched short videos of gesture and speech that conveyed semantically congruous and incongruous information. In half of the videos, participants were told that the two modalities were intentionally coupled (i.e., produced by the same communicator), and in the other half, they were told that the two modalities were not intentionally coupled (i.e., produced by different communicators). When participants knew that the same communicator produced the speech and gesture, there was a larger bi-lateral frontal and central N400 effect to words that were semantically incongruous versus congruous with gesture. However, when participants knew that different communicators produced the speech and gesture—that is, when gesture and speech were not intentionally meant to go together—the N400 effect was present only in right-hemisphere frontal regions. The results demonstrate that pragmatic knowledge about the intentional relationship between gesture and speech modulates controlled neural processes during the integration of the two modalities.

Keywords: Gesture; Speech; Language; Brain; Verbal; Nonverbal; Controlled; Semantic; Intention; N400

1. Introduction

Hand gestures, along with speech, are a natural part of communication.

Researchers have theorized that gesture and speech have an integrated relationship during language production (Kendon, 2004; McNeill, 1992) and have found that gestures influence language comprehension, using both behavioral techniques (for a review, see Goldin-Meadow, 2003) and elecrophysiological measures (Kelly, Kravitz, & Hopkins, 2004; Özyürek, Willems, Kita, & Hagoort, in press; Wu & Coulson, 2005). However, it is not known the degree to which people are in control, or are even aware, of this integration of gesture and speech. The present study explores whether instructing viewers about the intentional relationship between speech and gesture influences the neural integration of the two modalities during comprehension.

1.1. Gesture, speech and brain

The present study focuses on representational hand gestures. These gestures are hand configurations and movements that naturally, pervasively and spontaneously accompany speech in everyday communication. Their content typically conveys substantive and imagistic information about such things as object attributes, actions and spatial relationships. David McNeill theorizes that although representational gestures and speech convey information in quite different ways, the two channels are

☆ We would like to thank Colgate University's Psychology Department and Natural Science Division for emphasizing and supporting undergraduate research. We also thank Ron Crans for his technical support. Without his help, this project would not have been possible. In addition, we thank Peter Hagoort, Mandana Seyfeddinipur, David Wilkins, Susan Goldin-Meadow and Stephen Levinson for their helpful comments during various stages of this project.
* Corresponding author. Fax: +1 315 228 7942.
E-mail address: skelly@mail.colgate.edu (S.D. Kelly).

0093-934X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2006.07.008

conceptually linked during the earliest stages of language processing (McNeill, 1992). Much of the empirical research on this topic has focused on this relationship during language production, but psychologists have also made significant advances in the area of language comprehension (Beattie & Shovelton, 1999; Cassell, McNeill, & McCullough, 1999; Goldin-Meadow, Wein, & Chang, 1992; Kelly & Church, 1998). These studies strongly support the idea that gesture influences speech comprehension and, together with speech, may form an integrated system of communication.

As further evidence that gesture and speech are tightly integrated, researchers have theorized that spoken language systems emerged from primate gestural systems in our evolutionary past (Arbib & Rizzolatti, 1996; Corballis, 2003) and that gestures continue to play a significant role in present-day processing of speech (Kelly et al., 2002). This view is supported by recent research focusing on the neural level of analysis. Neuroimaging studies have found that brain regions that process speech also process actions made with the hand (Bonda, Petrides, Ostry, & Evans, 1996; Gallese, Keysers, & Rizzolatti, 2004; Nishitani, Schurmann, Katrin, & Hari, 2005; Puce & Perrett, 2003). For example, the superior temporal region in the left hemisphere is implicated not only in processing sound-based representations of speech (Hickok & Poeppel, 2000), but also goal-directed hand movements (Bonda et al., 1996). In addition, evidence from a transcranial magnetic stimulation (TMS) study demonstrated that when an experimenter magnetically disrupted the parts of the brain that control hand movements, speech comprehension also suffered (Floel, Ellger, & Breitenstein, 2003). This neural overlap of speech and hand regions strongly suggests that speech and gesture may be linked at some level in the brain.

This link is particularly evident in recent studies directly investigating the neural relationship between gesture and speech during language comprehension (Kelly et al., 2004; Özyürek et al., in press; Wu & Coulson, 2005). For example, Kelly and colleagues had participants watch videos of speech and gestural communication while event-related potentials (ERPs) recorded the semantic processing of the speech. In some videos, gestures conveyed congruent information to speech, and in others, gestures conveyed contrasting information to speech. The main result was that incongruous hand gestures produced a larger N400 component—reflecting semantic integration processes—to speech than congruous gestures, leading the researchers to conclude that gesture and speech are semantically integrated during language comprehension.

These studies provide ample evidence that gesture and speech may have an integrated relationship during language processing. An important next step is to better understand both the nature of and extent to this integrated relationship. For example, is this integration under some degree of conscious neurocognitive control, and if so, what communicative factors influence this integration?

1.2. Controlled processes and the intentional stance

Theories in cognitive psychology identify certain aspects of cognitive activity as *controlled* processes (Posner & Snyder, 1975; Schneider & Shiffrin, 1977). These types of cognitive processes are viewed as high-level, slow and under intentional control, in contrast to processes that are low-level, fast and obligatory. For example, strategically choosing how to interpret and act on some piece of communicative information is a type of controlled process. Put in these terms, how much control do we have over processing gestures that naturally accompany speech?

One possibility is that because gestures have such a fundamental relationship with speech (Goldin-Meadow, 2003; Kendon, 2004; McNeill, 1992), interlocutors may not be able to ignore gestural information during language processing. Indeed, researchers have argued that many types of non-linguistic communication (e.g., facial expression, body posture, tone of voice, and hand gesture) are processed automatically, immediately, and without conscious awareness (de Gelder, 2006; Kelly et al., 2004; Klucharev, Möttönen, & Sams, 2003; Lebib, Papo, de Bode, & Baudonniere, 2003; Pourtois, de Gelder, Vroomen, Rossion, & Crommelinck, 2000; van Wassenhove, Grant, & Poeppel, 2005; Winston, Strange, O'Doherty, & Dolan, 2002). For example, de Gelder and colleagues provided evidence for a sub-cortical neurobiological detection system that obligatorily and unconsciously processes emotional information conveyed through facial expressions and body language. A similar system may be in place for processing hand gestures. For example, research by Kelly et al. (2004) has shown that hand gestures that convey different representational information than the accompanying speech (e.g., gesturing the width of an object but verbally describing the height) impact very early sensory stages of speech processing.[1] One interpretation of this finding is that because sensory processing is thought to be obligatory and unconscious, integrating gesture and speech may not be under intentional control.

However, it is possible that integrating gesture and speech is not a completely obligatory process. For example, consider the following scenario: picture yourself in a café on a mobile phone talking to a friend while you watch two strangers engaged in an animated conversation that is rich with hand gestures. Although the visual information from the strangers' gestures might be quite salient and compelling, it should be easy to imagine separating the gestures you see from the words you hear on the telephone. In this way, integrating gesture and speech may be under some degree of neurocognitive control.

One way that people determine what to include with an interlocutor's speech is by making assessments of what that

[1] The present study, which uses a similar paradigm to the Kelly et al. (2004) study, uncovered comparable pre-semantic effects. However, because the current paper focuses only on controlled processes that are involved in semantic integration, we do not report these early sensory effects. These effects will be reported elsewhere in a different manuscript.

interlocutor intends to communicate. This sensitivity is what Daniel Dennett calls an *intentional stance*. Dennett argues that humans naturally believe that other people's behavior is intentional, and we cognitively use those beliefs to interpret language and other behaviors (Dennett, 1987). Indeed, human infants demonstrate this skill very early in development (Baldwin, 1993a, 1993b; Carpenter, Akhtar, & Tomasello, 1998; Meltzoff, 1995). For example, with regard to language learning, Baldwin (1993a) found that 19-month-old children used a communicator's physical manipulation of an object to infer a novel word referent only when the communicator looked at the manipulated object. Apparently, even very young children can ignore visual information during language learning when that information is not intentionally related to the speech it accompanies. These sorts of results generalize to hand gestures as well. We have long known that young language learners use gestures to infer communicative intent (Bates, 1976), and more recent work on adults has demonstrated that gestures have an intentional relationship with speech during language production (Bavelas, Kenwood, Johnson, & Phillips, 2002; Melinger & Levelt, 2004) and comprehension (Kelly, Barr, Church, & Lynch, 1999). For example, Melinger and Levelt (2004) observed that people often intentionally present information through gesture that is explicitly omitted from their speech during communication. On the comprehension side, Kelly et al. (1999) demonstrated that adults use speech and gesture together to infer the pragmatic intentions of a communicator when interpreting complex speech acts. These studies suggest that not only do speakers intentionally produce gestures in order to clarify speech, but interlocutors use gesture to clarify the intentions that underlie that speech.

There are even brain regions that appear dedicated to processing the intentions of hand movements (Allison, Puce, & McCarthy, 2000; Blakemore & Decety, 2001; Gallese et al., 2004; Grossman et al., 2000; Iacoboni et al., 2005; Pelphrey, Morris, & McCarthy, 2004; Rizzolatti & Craighero, 2004; Wohlschläger, Haggard, Gesierich, & Prinz, 2003). For example, Iacoboni et al. (2005) found that inferior frontal areas were more active when participants viewed reaching actions that were embedded in an intentional context than when viewing actions that were not. Although the majority of studies have focused on the intentional nature of non-communicative hand movements, recent work has generalized these findings and demonstrated that emblematic hand gestures, such as waving good-bye, also activate brain regions (e.g., the right superior temporal sulcus) that are sensitive to communicative intentions (Nakamura et al., 2004). Thus, when a communicator intends a gesture to accompany speech (as is typically the case: Bavelas et al., 2002; Melinger & Levelt, 2004), an interlocutor's sensitivity to intentions may facilitate integration of the two modalities into a coherent and holistic message. In the present study, we ask whether this ability to read intentions is under neurocognitive control and explore whether disrupting this ability—by instructing participants that gesture and speech do not belong together, as in the café example—modulates the controlled integration of gesture and speech during comprehension.

1.3. The present study

The present study investigates the neural integration of gesture and speech by using event-related potentials (ERPs). ERPs measure electrical brainwaves, produced by the post-synaptic discharge of large groups of neurons, that repeatedly follow (i.e., are "time locked" to) a particular stimulus of interest—in our case, spoken words. The brainwaves are averaged, and the end product is a series of peaks and valleys, called components, which correspond to different types of neruocognitive processes. As an index of gesture–speech integration, the study focuses on the classic N400 component. The N400 effect reflects the integration of a word into a previous semantic context (Kutas & Hillyard, 1980, 1984). For example, in response to the sentence, "The man liked cream and sugar in his socks," the word "socks" would produce a larger negative brainwave around 400 ms after presentation compared to a semantically appropriate word (such as "coffee"). In this way, the N400 component reflects neural processes that integrate speech into a previous context.

The N400 component is well suited for the present investigation for three main reasons. First, previous research using non-linguistic visual (including gestural) stimuli has uncovered N400 effects, though these effects had a more anterior scalp distribution than traditional text-based N400 effects, which are much more centrally distributed (Barrett & Rugg, 1990; Kelly et al., 2004; West & Holcomb, 2002; Wu & Coulson, 2005). For example, Wu and Coulson (2005) found that an incongruous gestural context produced a larger frontal and central N400 effect to words than a congruous gestural context. Based on this finding, the authors concluded that the N400 effect is sensitive not only to linguistic contexts, but to non-linguistic gestural contexts as well.

Second, there is an established body of research on the N400 component exploring the extent to which language comprehension is a controlled neurocognitive process (Brown & Hagoort, 1993; Chwilla, Brown, & Hagoort, 1995; Hill, Strube, Roesch-Ely, & Weisbrod, 2002; Holcomb, 1988; Kellenbach & Michie, 1996; Kiefer, 2002; Ruz, Madrid, Lupiáñez, & Tudela, 2003; Silva-Pereyra et al., 2003). For example, Kellenbach and Michie (1996) conducted a word pair priming study and found that N400 responses to targets were more robust when participants attended versus did not attend to the priming stimuli. They concluded from this finding that the sementic integration, as indexed by the N400 effect, is under some degree of neurocognitive control during word comprehension. Although there is debate over the extent to which the N400 effect reflects controlled processes (Balconi & Pozzoli, 2004; Heil, Rolke, & Peccinenda, 2004), these researchers all agree that the N400 component is a good candidate to investigate the issue. This is important because if gesture–speech integration is under some degree of neurocognitive control, the N400 component should be sensitive to participant's knowledge about the intentional relationship between the two modalities.

Third, the N400 component is sensitive to not just semantic contexts, but pragmatic contexts as well (Coulson, 2004; Hagoort, Hald, Bastiaansen, & Petersson, 2004; Van Berkum, Zwitserlood, Hagoort, & Brown, 2003). For example, Van Berkum and colleagues (2003) presented semantically appropriate sentences that were either congruous or incongruous with a wider discourse context. Sentences that were incongruous with the wider pragmatic context produced larger N400 effects than congruous sentences. Moreover, these discourse effects were identical to more traditional sentence-level N400 effects. Interestingly, this pragmatic influence generalizes to even wider discourse contexts, such as pre-existing beliefs about what people know to be true and false in the world (Hagoort et al., 2004). From these studies, it is clear that general pragmatic contexts play a powerful role in local semantic integration processes during language comprehension.

In the present experiment, we measured N400 responses to speech that was accompanied by semantically congruous and incongruous gestures. In addition, we manipulated participant's knowledge about whether gesture and speech were intended to go together by presenting speech–gesture pairs that were produced by the same communicator (an intentional relationship) or by different communicators (an unintentional relationship). The goal of this "intention" manipulation was to determine if general pragmatic knowledge—that is, knowledge about whether gesture and speech are meant to go together—modulates the integration of the two modalities. Based on these manipulations, we made two predictions. If gesture–speech integration is not under any neurocognitive control, pragmatic knowledge about the intentional relationship between gesture and speech should have no effect on integration, and the N400 effect should not differ across intentionality conditions. In contrast, if gesture–speech integration is under neurocognitive control, pragmatic knowledge about the intentional relationship between gesture and speech *should have* an effect on integration, and there should be an N400 effect only when there is an intentional, but not unintentional, relationship between the two modalities. Finally, based on previous ERP research using visually based stimuli (e.g., Özyürek et al., in press), we predicted that these effects would be most pronounced in more anterior electrode regions (frontal and central) of the scalp.

2. Method

2.1. Participants

Thirty-five right-handed (measured by the Edinburgh Handedness Inventory), Caucasian college undergraduates (15 males, 20 females; mean age: 20) participated for course credit. All participants signed an informed consent approved by the Institutional Review Board. A total of 31 participants contributed to the ERP analyses due to excessive artifacts in brainwave data from four participants.

Table 1
Summary of speech and gesture tokens in the Congruent and Incongruent conditions for the Intent and No Intent videos

	Speech			
	Tall	Thin	Short	Wide
Gesture				
Tall	15 Congruent	X	15 Incongruent	X
Thin	X	15 Congruent	X	15 Incongruent
Short	15 Incongruent	X	15 Congruent	X
Wide	X	15 Incongruent	X	15 Congruent

2.2. Materials

Participants watched digitized videos of two different people verbally and gesturally describing two objects: a tall, thin glass and a short, wide dish. In each clip, the person in the video uttered one of four speech tokens that corresponded to a salient dimension of the objects: *tall* and *thin* (the glass) and *short* and *wide* (the dish). These words were digitized and inserted into the video using iMovie software. To eliminate the contribution of lip and eye movements, only the torso of the communicators was shown in the videos. The people on the videos produced gestures relevant to the two objects, with different relationships between the speech and gesture. In the *Congruent* condition, the speech and gesture communicated the same information about the same object (e.g., saying tall and gesturing to the tallness of the glass). In the *Incongruent* condition, the speech and gesture presented different information about two different objects (e.g., saying tall and gesturing to the shortness of the dish). In addition to the "tall" speech stimuli, there were also three other speech stimuli: "thin," "short" and "wide." Within each relationship condition, each speech token was presented 15 times for a total of 60 tokens per relationship, for a total of 120 stimuli.[2] See Table 1.

In addition to the gesture–speech relationship condition, there was an intention condition that manipulated the intentional relationship between the gesture and speech. In the Intent video, the gesturer and speaker were the same person, but in the No Intent condition, the gesturer and speaker were two different people. The Intent condition was created by digitally inserting the speech of Person A and Person B into the video containing the gestures of Person A and Person B, respectively. In contrast, the No Intent condition was created by digitally inserting the speech of Person A and Person B into the video containing the gestures of Person B and Person A, respectively. In order to balance the presentation of the intention conditions, there were two different sets of videos. In one set (17 participants), the gestures of Person A were coupled with the speech of Person A

[2] The stimulus tapes included a third condition in which the gestures conveyed complementary information to speech (e.g., gesturing to the *thinness* of the glass and saying *tall*.) These stimuli produced early ERP effects (see the first footnote) but did not produce later semantic effects, such as the N400 effect. Because the present study focuses exclusively on the N400 effect, these stimuli are not included in the present analyses.

(Intent condition), and the gestures of Person B were coupled with the speech of Person A (No Intent condition). In the second set (14 participants), the gestures of Person B were coupled with the speech of Person B (Intent condition), and the gestures of Person A were coupled with the speech of Person B (No Intent condition). See Fig. 1 for an example of all the treatment combinations focusing on the "tall" speech token. Note that the videos were constructed so that the Intent and No Intent conditions were balanced across Persons A and B, so that any differences between the intent conditions must be due to knowledge about the intentional relationship between the speech and gesture of the communicator, and not to the physical attributes of the gesture and speech.

Each audiovisual segment was 3 s in duration, with the first portion showing a static image of the actor sitting behind the objects facing the camera. In all videos, the speech onset occurred at 2 s. All of the ERP data were time-locked to speech onset. The preparation of the gestures (i.e., the hand movement that preceded the actual content of the gesture) began an average of 16.25 frames (542 ms) prior to speech onset, and the stroke (i.e., the onset of the meaningful portion of the gesture) occurred at speech onset. The gesture was held in position near the object until the end of the video. The variable inter-stimulus interval was 1.0–1.5 s. Each video was 13 min in length.

3. Procedure

After participants were fitted with the ERP net (see below), the experimenter explained the intention manipulation. Participants first watched two short videos (head and torso) of two people, Persons A and B, introducing themselves. The purpose of this short introduction was to demonstrate which voice went with which body. After this short video introduction, the experimenter explained that participants would watch a series of short video clips containing the speech and gesture of Persons A and B. However, half of the subsequent video clips were digitally edited such that the speech and gesture did not belong together. That is, one video contained gestures that were meant to accompany the speech (the Intent video), but the other video contained gestures that were not meant to accompany the speech (the No Intent video). The experimenter then showed two sample video clips, with the head and mouth visible, in which the speech and gesture were not meant to accompany one another (e.g., showing a brief video of Person A gesturing with an over-dub of Person B speaking, or vice versa). The head and mouth were visible to highlight the intention manipulation—it was clear that the words that participants heard did not match the lip movements that they saw on the video. The experimenter explained that participants would view only the torso of the gesturers in the experimental trials, but they were to keep in mind that in the No Intent video, the speech and gesture were not meant to accompany one another.

Following this, participants were told that their task during the experimental trials was to attend to the speech in the video clips. Specifically, participants were to press one key on a button box if the speech referred to the glass and another key if the speech referred to the dish. Participants were instructed to produce a response as soon as they heard the speech rather than wait for a prompt. A computer recorded these responses, and their latencies were used in the reaction time analyses. In addition, the computer simultaneously recorded brainwave data to the same responses. Note that the task does not require any attention to the gestural information. Participants then watched either the Intention or No Intention video first (the order was counterbalanced), and in this way, the Intent conditions were blocked rather than mixed during stimulus presentation. Importantly, to balance the intention conditions equally across Persons A and B, 17 participants watched Set 1 and 14 participants watched Set 2. Refer to Fig. 1.

3.1. ERP set up and analysis

Participants were fitted with a 128-electrode Geodesic ERP net. The EEG was sampled at 250 Hz using a band pass filter of 0.1–30 Hz. The impedances were kept below 40 kohm (the Geonet system uses high-impedance amplifiers).

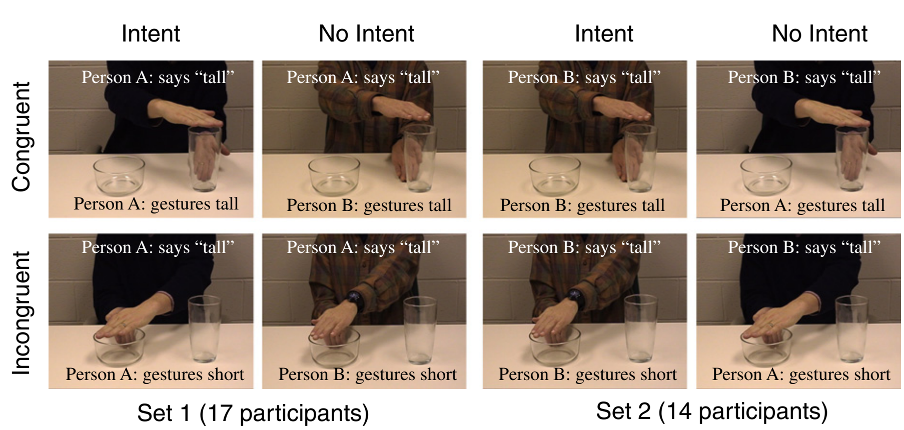

Fig. 1. Still frames of the Congruent and Incongruent stimuli for the Intent and No Intent videos for the *tall* verbal stimuli. Person A is in the solid shirt, and Person B is in the plaid shirt. The Intent and No Intent videos contained 120 stimuli each (described in Table 1).

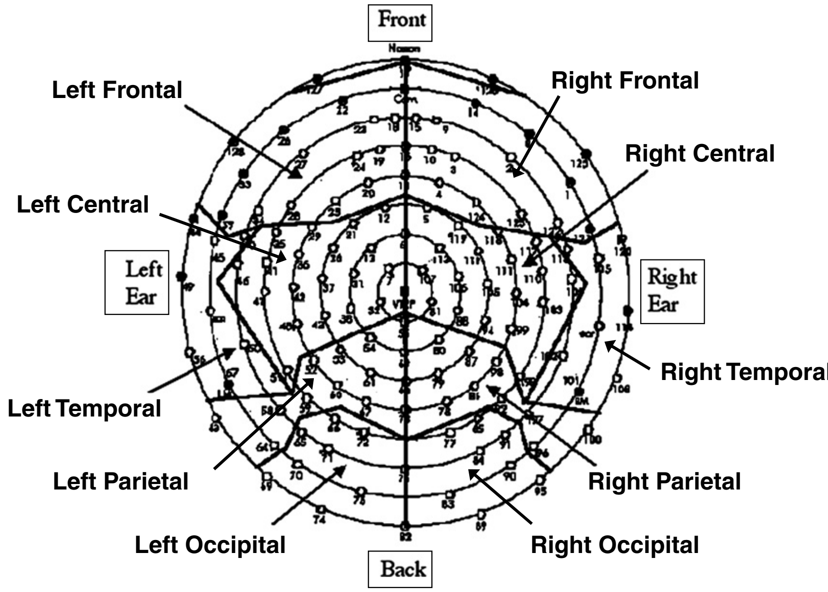

Fig. 2. Ten electrode clusters for the 128 Geodesic electrode net. For more on the rationale for the clusters, see Kelly et al. (2004).

Eye artifacts were monitored with 4 EOG electrodes. All artifacts were corrected offline. The ERPs were vertex referenced for recording and linked-mastoid referenced for analysis and presentation (to facilitate comparison to other N400 findings in the literature). The data were electronically scanned for movement artifacts, and all bad trials were rejected. On average, there were 6.67 ($SD = 4.41$) rejected trials in the Intent Congruent condition, 6.73 ($SD = 4.13$) in the Intent Incongruent condition, 6.07 ($SD = 3.54$) in the No Intent Congruent condition, and 5.77 ($SD = 3.12$) in the No Intent Incongruent condition.

The behavioral analysis was based on a 2 (Intent, No Intent) by 2 (Congruent, Incongruent) repeated measures design with response latencies as the dependent measure. The ERP analysis was based on a 2 (Intent, No Intent) by 2 (Congruent, Incongruent) by 5 (Central, Frontal, Occipital, Parietal, Temporal Electrode Region) by 2 (Right, Left Hemisphere) repeated measures design. The analyses focused on the N400 effect, which was created by averaging the ERP amplitude from a window of 350–700 ms post-speech stimulus. The electrode manipulation refers to the location of clusters of electrodes on the scalp. Based on previous studies, the 128 electrodes were broken up into five clusters of channels that corresponded roughly to basic anatomical structures of the brain. Refer to Fig. 2 for a diagram of the clusters for the 128 electrodes (for more on the clusters, see Kelly et al., 2004).

All repeated measures analyses were adjusted for the problem of sphericity by using a Greenhouse–Geisser correction. Because the present study advanced a sub-set of particular a priori hypotheses about the Intent and No Intent conditions, planned *t* tests compared Congruent and Incongruent effects at left and right hemisphere frontal and central electrode sites within each of the Intent conditions.[3] Although all contrasts were planned and orthogonal, a Dunn–Sidak correction was applied within each of the two Intent conditions in order to reduce the likelihood of producing Type I errors.

4. Results

4.1. Behavioral data

Participants were near ceiling with the accuracy scores (approximately 99% accurate), and there were no differences among groups. The analysis for the reaction time scores focused on correct responses from the original 35 particpants. For the reaction time data, there was not a significant main effect of Intention, $F(1, 34) = 0.27$, *ns*, or an Intention by Congruence interaction, $F(1,34) = 3.84$, ns, but there was a significant main effect of Congruence, $F(1, 34) = 48.32$, $p < .001$. From inspection of Table 2, it is clear that Incongruent stimuli were slower than Congruent stimuli in both the Intent and No Intent video conditions.

Recall that in order to balance the Intent conditions equally across Persons A and B in the video, two groups of participants watched two different sets of stimuli, with 17 participants watching Set 1 and 14 participants watching Set 2 (refer to Fig. 1). Because the data were collected on these two separate sets of participants, we included Set as a

[3] Although there are several general questions that we could address by exploring the omnibus difference among all means of the 4-way design in the present study, we are concerned with a specific, theoretically driven sub-set of the variables for the purposes of this paper. Because we had a priori predictions about this sub-set of variables, and because we wanted to reduce the likelihood of committing Type II errors, we followed the recommendations of Kirk (1995) and Abelson (1995) and presented planned orthogonal *t tests* to explore N400 differences between Congruent and Incongruent stimuli for the Intent and No Intent conditions.

Table 2
Reaction time and Standard Deviations for Congruent and Incongruent stimuli for the Intent and No Intent videos

	Intent video		No intent video	
	Mean	*SD*	Mean	*SD*
Congruent	642 ms	142 ms	657 ms	139 ms
Incongruent	689 ms	140 ms	691 ms	130 ms

between-subjects factor in our ANOVA on the RT scores. The only significant Set effect was a Set by Gesture interaction, $F(1, 34) = 15.66$, $p < .001$. This interaction was driven by the participants producing a greater difference between Congruent and Incongruent conditions in Set 2 ($M = 61$ ms, $SD = 35$ ms) compared to Set 1 ($M = 23$ ms, $SD = 22$ ms), $t(34) = 3.96$, $p < .001$.

4.2. ERP data

The ANOVA did not reveal a significant effect of Intention, but there was a significant main effect of Gesture, $F(1, 30) = 5.41$, $p = .027$, and a significant Gesture by Electrode interaction, $F(4, 120) = 12.81$, $p < .001$. As predicted, this effect was driven by two electrode regions: Incongruent stimuli produced a larger N400 effect than Congruent stimuli in bi-lateral central, $tDS(2, 30) = 4.28$, $p < .001$, and frontal regions, $tDS(2, 30) = 3.92$, $p < .001$.

The next analyses focused on the planned contrasts comparing Congruent and Incongruent stimuli within the different Intent video conditions. Although the 4-way interaction of Intent by Gesture by Electrode by Hemisphere was not significant, $F(4, 120) = .23$, ns, planned Dunn–Sidak contrasts revealed an interesting pattern between the Congruent and Incongruent conditions for the Intent and No Intent videos in left and right central and frontal electrode regions. In the Intent condition, Incongruent stimuli produced a larger N400 effect than Congruent stimuli in central sites for the left hemisphere, $tDS(4, 30) = 3.07$, $p < .05$, and right hemisphere, $tDS(4, 30) = 3.16$, $p < .05$. This pattern also held in frontal regions, with Incongruent stimuli producing a larger N400 effect than Congruent stimuli in both the left, $tDS(4, 30) = 3.23$, $p < .05$, and right hemispheres, $tDS(4, 30) = 3.79$, $p < .01$.

A different pattern was evident in the No Intent condition. In this condition, Incongruent stimuli did not produce a larger N400 effect than Congruent stimuli in central regions in the left hemisphere, $tDS\ (4, 30) = 1.57$, ns, or right hemisphere, $tDS\ (4, 30) = 1.91$, ns. In addition, Incongruent stimuli were no different than Congruent stimuli in left frontal regions, $tDS\ (4, 30) = 0.92$, ns. However, in right frontal regions, Incongruent stimuli did produce a larger N400 effect than Congruent stimuli, $tDS\ (4, 30) = 2.62$, $p < .05$. Refer to Fig. 3 for brainwave data for each of our ten electrode clusters and Fig. 4 for the means and standard errors for the windowed N400 effects at frontal and central electrode regions. Fig. 5 presents the data in a conventional mastoid-referenced 10–20 array to illustrate the effects at particular electrode sites.

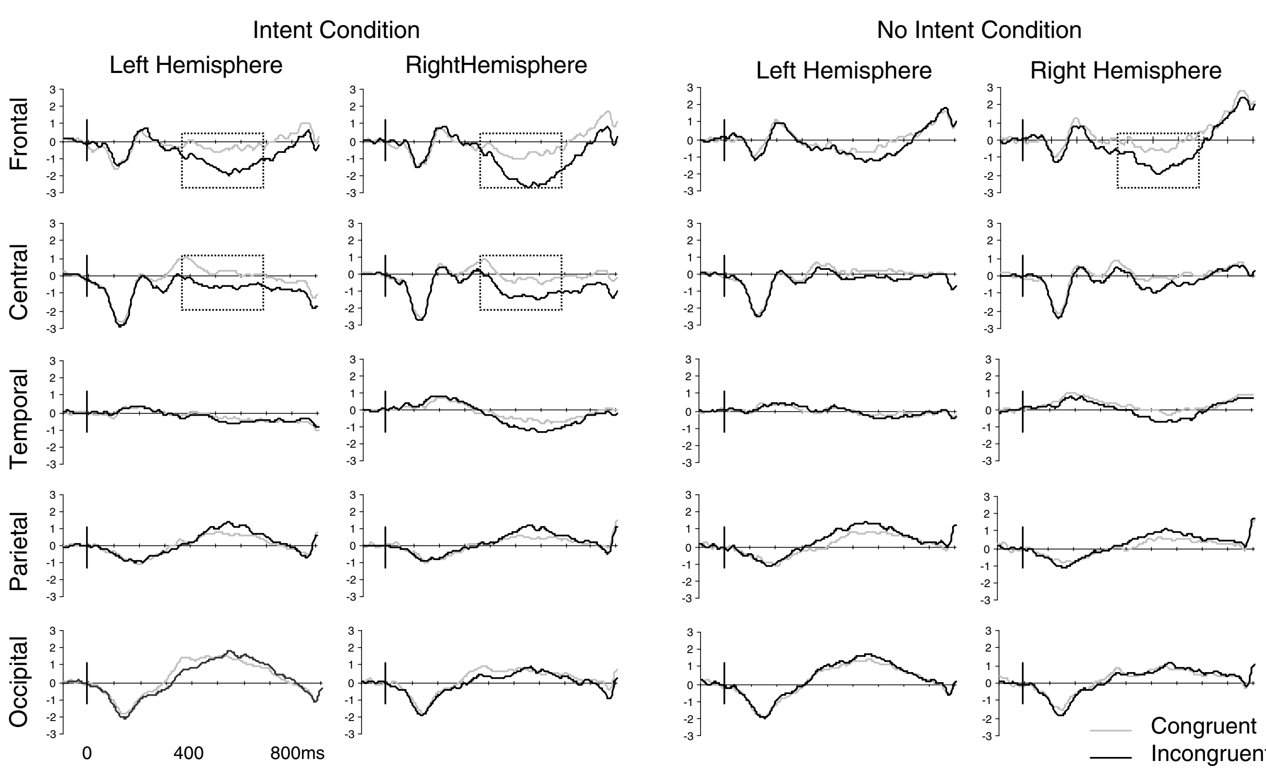

Fig. 3. ERP responses were time-locked to the speech in the Congruent and Incongruent stimuli at each of the 10 electrode regions for the Intent and No Intent video conditions. The windowed regions (350–700 ms) show the significant N400 differences between the Congruent and Incongruent conditions. Time zero corresponds to the speech onset, and the brainwaves extend to 900 ms.

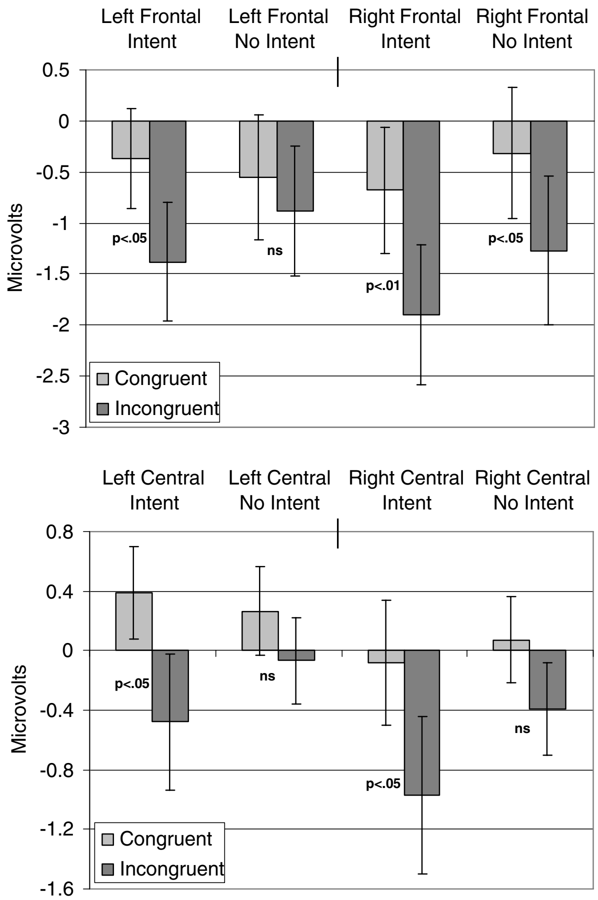

Fig. 4. The means and standard errors of the windowed N400 effects (350–700 ms) are presented for the Intent and No Intent video conditions at frontal and central electrode regions.

Finally, in order to determine whether Set 1 and 2 produced different effects for each of the conditions, we entered Set as a between-subjects factor in our ANOVA. Unlike the behavioral results, the Set condition did not produce any additional significant effects.

5. Discussion

The present study replicated previous research demonstrating ERP differences to speech when it is accompanied by semantically incongruent versus congruent gestures (Kelly et al., 2004; Özyürek et al., in press; Wu & Coulson, 2005). Specifically, the Incongruent stimuli produced a larger N400 effect than Congruent stimuli in bi-lateral central and frontal electrode regions, as predicted. However, this N400 effect was present in left and right central and frontal sites for the Intent video condition, but it was present only in right frontal sites for the No Intent video condition. These findings support the second prediction advanced in this paper: pragmatic knowledge about the intentional relationship between gesture and speech does affect the integration of gesture and speech at comprehension.

Before we explore the ERP results in more detail, the behavioral results merit attention. One possible explanation for the ERP findings is that participants simply did not pay attention to the gestures in the No Intent condition. However, the reaction time data do not support this alternative explanation. Recall that participants produced slower reaction times for Incongruent compared to Congruent stimuli in both the Intent and No Intent video conditions (Table 2). This suggests that participants must have attended to the incongruent gestures in both videos.

Although the reaction time differences between Congruent and Incongruent stimuli were similar for both Intent conditions, the ERP data tell a different story. The planned contrasts revealed that in the Intent condition, the N400 component to Incongruent stimuli was present in bi-lateral central and frontal regions, whereas in the No Intent condition, it was present only in the right hemisphere frontal region. This inconsistency between the behavioral and electrophysiological findings may appear odd, but previous research on the N400 effect using a semantic priming paradigm yielded similar results (Brown & Hagoort, 1993; Holcomb, 1993; Ruz et al., 2003). For example, Brown and Hagoort (1993) found that semantically incongruent versus congruent primes produced reaction time differences to targets under conditions in which primes were masked and unmasked. However, only incongruent unmasked primes produced a corresponding N400 effect, whereas incongruent masked primes did not. Although the present study is different because it did not measure unconscious semantic processing, there are similarities in that both studies uncovered ERP effects that were inconsistent with the behavioral effects. It is not clear whether the two different dependent measures in the present study yielded different results simply because reaction time measures are not as sensitive as electrophysiological measures, or whether the two measures actually tapped into two different aspects of gesture–speech integration. For example, one could speculate that the reaction time results reflect an early and automatic influence of gesture on speech processing (due to sensory priming and automatic spread of activation), whereas the ERP results reflect a later and more controlled type of processing (due to semantic integration processes). The current study was not designed to validate this possibility, and future studies should investigate this interesting issue more directly.

The fact that the No Intent manipulation did significantly limit the integration between gesture and speech suggests that the process of integration may be under some degree of neurocognitive control. This finding fits well with previous studies that have shown that general discourse information can influence the N400 component to language stimuli (Coulson, 2004; Hagoort et al., 2004; Van Berkum et al., 2003). For example, Van Berkum et al. (2003) presented identical sentences in wider discourse contexts and found that semantically appropriate, but pragmatically anomalous, target words (i.e., words that were ironic in context) produced a larger N400 effect than pragmatically expected words (i.e., words that were literal in context). Other studies have demonstrated that simply instructing people to process the same information differently can sig-

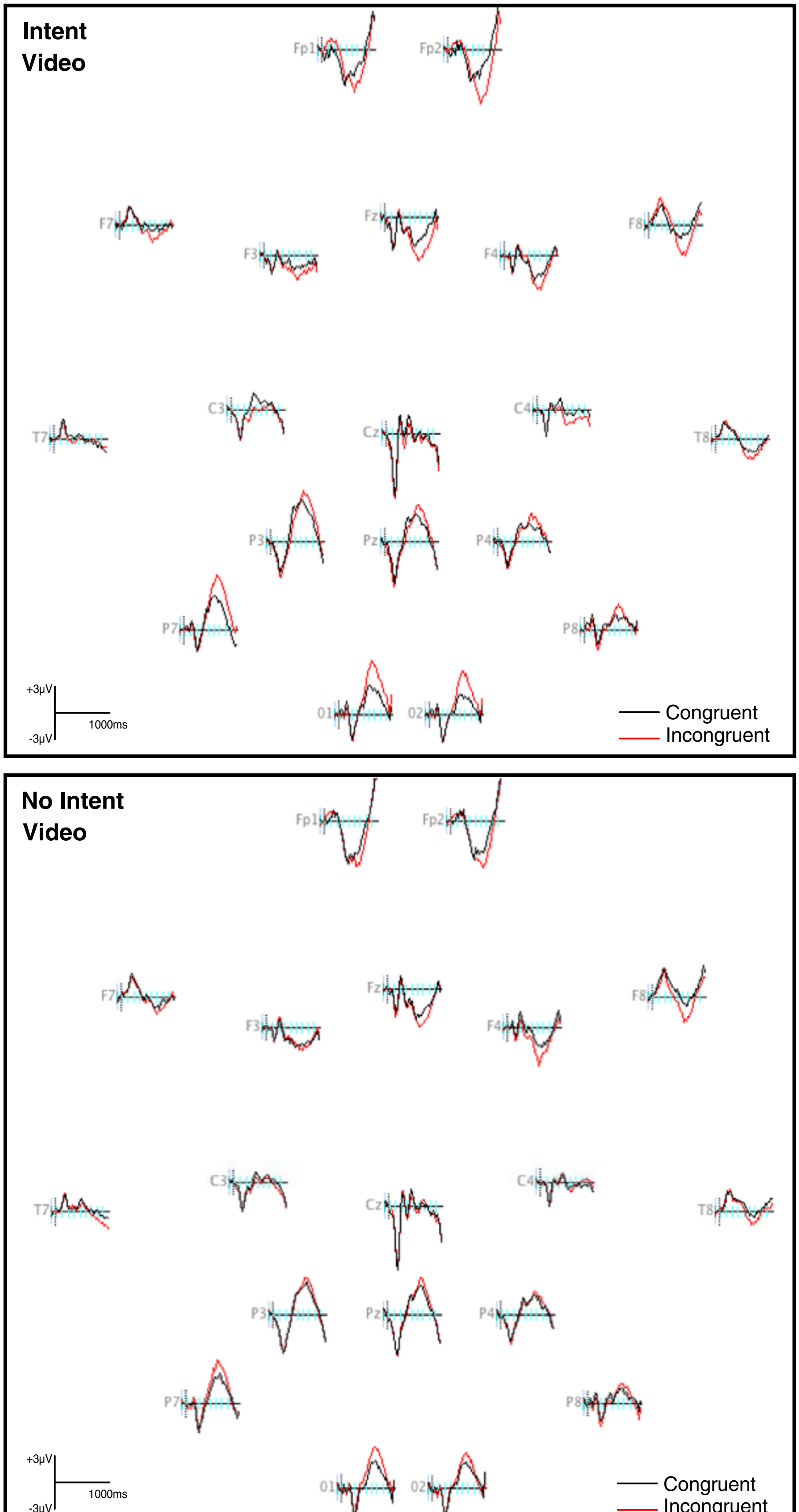

Fig. 5. Mastoid-referenced ERP data at individual electrode sites according to the 10–20 international system.

nificantly influence the neural processing of a stimulus (Sabbagh, Moulson, & Harkness, 2004; Winston et al., 2002). For example, Sabbagh and colleagues had people analyze identical visual stimuli, pictures of human eyes, and found a larger anterior negativity (N270–400) when people judged the emotions of the eyes versus the sex of the eyes. Similarly, the present study presented identical stimuli to participants, videos of speech and gesture, and manipulated knowledge about whether gesture and speech were meant to go together. This knowledge may have served as a sort of general discourse context that differentially influenced semantic integration (reflected by the N400 effect) of the two modalities.[4]

This interpretation of the data fits well with more general philosophical views on how judgments of intentionality are crucial to human communication (Dennett, 1987; Grice, 1957). For example, applying Dennett's notion of the "intentional stance" to the present results, it is possible that interlocutors in everyday discourse may naturally (and safely) assume that gestures that accompany speech are intended to go with that speech, and under normal circumstances, they may integrate speech and gesture by default. However, in a typical circumstances in which the intentional pairing between speech and gesture is disrupted (as in the present experiment), people may override this stance and "choose" not to integrate gesture and speech. In this way, people's intentional stance toward the relationship between speech and gesture may serve as a general context under neurocognitive control that, in typical situations, empowers gesture to influence speech comprehension.

Although it is clear that our Intent manipulation influenced gesture–speech integration in the present experiment, the exact nature of this effect deserves more attention. Note that our N400 effect had a generally anterior distribution. Although previous research on the N400 component using strictly text-based stimuli has traditionally found a centro-parietal distribution (Kutas & Hillyard, 1980), a more anterior distribution of the N400 effect is consistent with research using non-text based stimuli, such as pictures or videos (Barrett & Rugg, 1990; Sitnikova, Kuperberg, & Holcomb, 2003; West & Holcomb, 2002). On the surface, the present results may appear to merely replicate previous image-based N400 findings. However, the present study differs from previous studies in at least one important way. Previous research on picture processing has used linguistic or imagistic information to set up a semantic context (a prime) and taken the ERPs to an image-based target, whereas the present study set up a visual context and measured ERPs to an accompanying speech target. Considering this difference, it is interesting that the anterior scalp distribution of the N400 effect in the present study was more similar to the distribution of the N400 effect to images rather than to other speech stimuli (i.e., the more classic centro-parietal pattern). In this way, it is possible that a visual context may cause speech to be processed in a more imagistic, rather than linguistic, fashion. This finding is interesting because it suggests that measuring ERPs to linguistic stimuli in more complex and embodied contexts—dynamic multimodal videos of human communicators versus isolated words presented on a computer screen—substantially shifts the scalp topography of the N400 component.

Another important aspect of the results is that there were interesting hemispheric differences in the distribution of the N400 component. Recall that the Incongruent condition produced a pronounced N400 component in bilateral frontal and central electrode sites for the Intent video, but it produced only a right-lateralized frontal effect in the No Intent video. Although one must be careful when speculating about hemisphere effects based on ERP scalp topographies, it is interesting that this pattern fits well with a wide range of research employing diverse methods demonstrating that the right hemisphere is more sensitive to indirect and distant semantic relationships than the left hemisphere (Beeman et al., 1994; Chiarello, 1988; Coulson & Wu, 2005; Federmeier & Kutas, 1999; Hagoort, Brown, & Swaab, 1996; Kiefer, Weisbrod, Kern, Maier, & Spitzer, 1998). For example, Kiefer et al. (1998) conducted a semantic priming ERP study and found that the right hemisphere distinguished between indirectly related semantic pairs (lemon–sweet) and non-related semantic pairs (leaf–car), but it did not differentiate between indirectly (lemon–sweet) and directly related pairs (hen–egg). In contrast, the left hemisphere distinguished between directly related and non-related pairs, but not between indirectly related and non-related pairs. This finding suggests that the right hemisphere may be specialized to make more elaborate semantic links among concepts than the left hemisphere.

If one views our No Intent manipulation as weakening the semantic link between speech and gesture, it follows that the left hemisphere (which is less sensitive to semantic distance) would lose sensitivity to gesture–speech incongruencies, whereas the right hemisphere (which is more sensitive to semantic distance) would continue to integrate incongruent gestures and speech. This possibility is consistent with a laterality model by Federmeier and Kutas (1999) stating that the right hemisphere is "integrative" in that it coarsely (and perhaps automatically) compares incoming speech to the general context, whereas the left hemisphere is "predictive" in that it more strategically focuses on the semantic meaning of a message to build expectations about upcoming linguistic input. Relating this model to the present results, the No Intent condition may have disrupted the gestures from producing semantic expectations in the left hemisphere and thus weakened the predictive power and effect of those gestures on subsequent speech processing.

[4] It is difficult to determine whether our manipulation successfully changed participants' *beliefs* about the intentional relationship between speech and gesture, or whether it simply made them aware of what the experimenters wanted them to *know* about the speech and gesture. We are currently conducting a study that attempts to pull these two issues apart.

A different way to view the laterality results is to consider the literature on hemispheric differences in processing the intentions and mental states of others (Grèzes, Frith, & Passingham, 2004; Iacoboni et al., 2005; Nakamura et al., 2004; Pelphrey et al., 2004; Sabbagh et al., 2004; Winston et al., 2002). For example, Iacoboni and colleagues (2005) found that right hemisphere inferior frontal regions were particularly sensitive to hand grasping behaviors that were embedded in intentional contexts. Importantly, the authors found that this right hemisphere effect did not differ when participants were and were not explicitly instructed to analyze the intentions of the hand grasps. However, this instruction not to analyze the intentions of the hand grasps did reduce activity in left mesial frontal and cingulate cortices. Thus, explicit instructions regarding attention to hand actions impacted left more than right hemisphere processing. Relating this finding to the current study, it is possible that explicit knowledge about the lack of an intentional relationship between speech and gesture in the No Intent condition disrupted left hemisphere processing to a greater extent than right hemispheric processing. This finding suggests that processing gesture and speech may be under different degrees of neurocognitive control in the left and right hemispheres.

In conclusion, the findings of the present study suggest that pragmatic knowledge about the intentional relationship between speech and gesture modulates hemispheric integration of the two modalities during language comprehension. Although the results are generally consistent with the view that gesture–speech integration may involve controlled processes, future research will need to address important remaining questions: Does pragmatic information about the intentional relationship between gesture and speech have any influence on earlier pre-semantic processes involved in gesture–speech integration? Do the present effects generalize to more natural communicative situations in which intentions are inferred rather than dictated? To what extent do actual intentions of a communicator versus an interlocutor's knowledge about a communicator's intentions influence gesture–speech comprehension? Answers to these questions will be necessary in order to further our knowledge on how pragmatic information about mental states and intentionality contributes to the neural integration of verbal and gestural information during language comprehension.

References

Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Lawrence Erlbaum.

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Science, 4*, 267–278.

Arbib, M. A., & Rizzolatti, G. (1996). Neural expectations: a possible evolutionary path from manual skills to language. *Communication and Cognition, 29*, 393–424.

Balconi, M., & Pozzoli, U. (2004). N400 and P600 or the role of the ERP correlates in sentence comprehension: some applications to the Italian language. *The Journal of General Psychology, 131*, 268–302.

Baldwin, D. A. (1993a). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology, 29*, 832–843.

Baldwin, D. A. (1993b). Infant's ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*, 395–418.

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition, 14*, 201–212.

Bates, E. (1976). *Language and context: The acquisition of pragmatics*. New York: Academic Press.

Bavelas, J., Kenwood, C., Johnson, T., & Phillips, B. (2002). An experimental study of when and how speakers use gestures to communicate. *Gesture, 2*, 1–17.

Beattie, G., & Shovelton, H. (1999). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica, 123*, 1–30.

Beeman, M., Friedman, R. B., Grafman, J., Perex, E., Diamond, S., & Lindsay, M. B. (1994). Summation priming and coarse semantic coding in the right hemisphere. *Journal of Cognitive Neuroscience, 6*, 26–45.

Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews: Neuroscience, 2*, 561–567.

Bonda, E., Petrides, M., Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience, 16*, 3737–3744.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5*, 34–44.

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior & Development, 21*, 315–330.

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

Chiarello, C. (1988). Semantic priming in the intact brain: Separate roles for the left and right hemispheres? In C. Chiarello (Ed.), *Right hemisphere contributions to lexical semantics* (pp. 59–69). New York: Springer.

Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of level of processing. *Psychophysiology, 32*, 274–285.

Corballis, M. C. (2003). *From hand to mouth: The origins of language*. Princeton, NJ: Princeton University Press.

Coulson, S. (2004). Electrophysiology and pragmatic language comprehension. In I. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 187–206). San Diego: Palgrave Macmillan.

Coulson, S., & Wu, Y. C. (2005). Right hemisphere activation of joke-related information: An event-related brain potential study. *Journal of Cognitive Neuroscience, 17*, 494–506.

de Gelder, B. (2006). Towards the neurobiology of emotional body language. *Nature Reviews: Neuroscience, 7*, 242–249.

Dennett, D. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Federmeier, K., & Kutas, M. (1999). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research, 8*, 373–392.

Floel, A., Ellger, T., & Breitenstein, C. (2003). Language perception activates the hand motor cortex: Implications for motor theories of speech perception. *European Journal of Neuroscience, 18*, 704–708.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Science, 8*, 386–403.

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

Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge through gesture: Using children's hands to read their minds. *Cognition and Instruction, 9*, 201–219.

Grèzes, J., Frith, C., & Passingham, R. E. (2004). Brain mechanisms for inferring deceit in the actions of others. *Journal of Neuroscience, 24*, 5500–5505.

Grice, P. (1957). *Meaning. Philosophical Review, 66*, 377–388.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., et al. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience, 12*, 711–720.

Hagoort, P., Brown, C., & Swaab, T. (1996). Lexical-semantic event-related potential effects in patients with left hemisphere lesions and aphasia, and patients with right hemisphere lesions without aphasia. *Brain, 199*, 627–649.

Hagoort, P., Hald, L., Bastiaansen, M. C. M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*, 438–440.

Heil, M., Rolke, B., & Peccinenda, A. (2004). Automatic semantic activation is no myth: Semantic context effects on the N400 in the letter search task in the absence of RT effects. *Psychological Science, 15*, 852–857.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Science, 4*, 131–138.

Hill, H., Strube, M., Roesch-Ely, D., & Weisbrod, M. (2002). Automatic vs. controlled processes in semantic priming - differentiation by event related potentials. *International Journal of Psychophysiology, 44*, 197–218.

Holcomb, P. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language, 35*, 66–85.

Holcomb, P. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology, 30*, 47–61.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping intentions of others with one's own mirror neuron system. *PloS Biology, 3*, 529–535.

Kellenbach, M. I., & Michie, P. T. (1996). Modulation of event-related potentials by semantic priming: Effects of color-cued selective attention. *Journal of Cognitive Neuroscience, 8*, 155–173.

Kelly, S. D., & Church, R. B. (1998). A comparison between children's and adults' ability to detect children's representational gestures. *Child Development, 69*, 85–93.

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

Kelly, S. D., Iverson, J., Terranova, J., Niego, J., Hopkins, M., & Goldsmith, L. (2002). Putting language back in the body: Speech and gesture on three timeframes. *Developmental Neuropsychology, 22*, 323–349.

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

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

Kiefer, M., Weisbrod, M., Kern, I., Maier, S., & Spitzer, M. (1998). Right hemisphere activation during indirect semantic priming: Evidence from event-related potentials. *Brain and Language, 64*, 377–408.

Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences*. Pacific Grove, CA: Brooks/Cole Publishing Company.

Klucharev, V., Möttönen, R., & Sams, M. (2003). Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cognitive Brain Research, 18*, 65–75.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentialsreflect semantic incongruity. *Science, 207*, 203–204.

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

Lebib, R., Papo, D., de Bode, S., & Baudonniere, P. M. (2003). Evidence of a visual–auditory cross-modal sensory gating phenomenon as reflected by the human P50 event-related potential modulation. *Neuroscience Letters, 341*, 185–188.

McNeill, D. (1992). *Hand and mind: What gesture reveals about thoughts*. Chicago, IL: University of Chicago Press.

Melinger, A., & Levelt, W. (2004). Gesture and the communicative intention of the speaker. *Gesture, 4*, 119–141.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 838–850.

Nakamura, A., Maess, B., Knösche, T. R., Gunter, T. C., Bach, P., & Friederici, A. D. (2004). Cooperation of different neuronal systems during hand sign recognition. *Neuroimage, 23*, 25–34.

Nishitani, N., Schurmann, M., Katrin, A., & Hari, R. (2005). Broca's region: From action to language. *Physiology, 20*, 60–69.

Özyürek, A., Willems, R., Kita, S., & Hagoort, P. (in press). On-line integration of semantic information from speech and gesture: Insights from event-related potentials. *Journal of Cognitive Neuroscience*.

Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2004). Grasping the intentions of others: The perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of Cognitive Neuroscience, 16*, 1706–1716.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., & Crommelinck, M. (2000). The time-course of intermodal binding between hearing and seeing affective information. *Neuroreport, 11*, 1329–1333.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55–85). Hillsdale, NJ: Lawrence Erlbaum.

Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical transactions of the Royal Society of London, 385*, 435–445.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169–192.

Ruz, M., Madrid, E., Lupiánez, J., & Tudela, P. (2003). High density ERP indices of conscious and unconscious semantic priming. *Cognitive Brain Research, 17*, 719–731.

Sabbagh, M. A., Moulson, M. C., & Harkness, K. L. (2004). Neural correlates of mental state decoding in human adults: An event-related potential study. *Journal of Cognitive Neuroscience, 16*, 415–426.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: Detection, search, and attention. *Psychological Review, 84*, 1–66.

Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology, 40*, 160–164.

Van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research, 17*, 701–718.

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

West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research, 13*, 363–375.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience, 5*, 277–283.

Wohlschläger, A., Haggard, P., Gesierich, B., & Prinz, W. (2003). The perceived onset time of self- and other-generated actions. *Psychological Science, 14*, 586–591.

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology, 42*, 654–667.