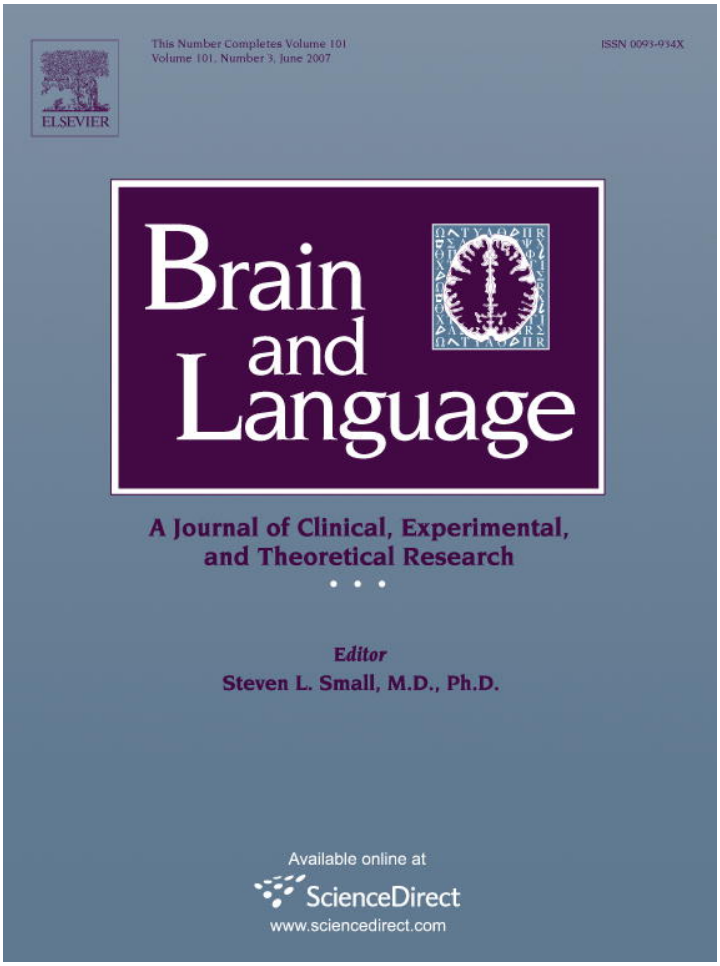

Available online at www.sciencedirect.com
ScienceDirect
Brain and Language 101 (2007) 234–245
Brain and Language
www.elsevier.com/locate/b&l

How iconic gestures enhance communication: An ERP study

Ying Choon Wu, Seana Coulson *

Cognitive Science Department 0515, 9500 Gilman Drive, La Jolla, CA 92093-0515, USA

Accepted 5 December 2006
Available online 12 January 2007

Abstract

EEG was recorded as adults watched short segments of spontaneous discourse in which the speaker's gestures and utterances contained complementary information. Videos were followed by one of four types of picture probes: cross-modal related probes were congruent with both speech and gestures; speech-only related probes were congruent with information in the speech, but not the gesture; and two sorts of unrelated probes were created by pairing each related probe with a different discourse prime. Event-related potentials (ERPs) elicited by picture probes were measured within the time windows of the N300 (250–350 ms post-stimulus) and N400 (350–550 ms post-stimulus). Cross-modal related probes elicited smaller N300 and N400 than speech-only related ones, indicating that pictures were easier to interpret when they corresponded with gestures. N300 and N400 effects were not due to differences in the visual complexity of each probe type, since the same cross-modal and speech-only picture probes elicited N300 and N400 with similar amplitudes when they appeared as unrelated items. These findings extend previous research on gesture comprehension by revealing how iconic co-speech gestures modulate conceptualization, enabling listeners to better represent visuo-spatial aspects of the speaker's meaning.

Keywords: Gesture; N400; N300; Semantic integration; Language comprehension; Object recognition; Conceptual integration; Embodiment; ERP; Meaning; Simulation

1. Introduction

Co-speech gestures provide a channel for speakers to express additional information related to their communicative intent. While uttering, "It's actually a double door," for example, a speaker may indicate the shape of a Dutch rather than French style door with the configuration of his hands (see Fig. 1). A number of behavioral studies suggest that gestures such as these play a beneficial role in communication. Listeners rely on speakers' gestures to disambiguate communicative intent in cases where understanding may be impeded—due to noise in the speech signal, for example (Rogers, 1978; Thompson & Massaro, 1986, 1994), or due to additional inferential processing engendered by indirect requests (Kelly, 2001; Kelly, Barr, Church, & Lynch, 1999). Listeners also exhibit a more accurate understanding of instructions and narratives when the speaker's accompanying gestures are visible (Beattie & Shovelton, 1999b, 2002; Graham & Argyle, 1975; Morford & Goldin-Meadow, 1992; Singer & Goldin-Meadow, 2005; Valenzeno, Alibali, & Klatzky, 2003). However, see Krauss, Dushay, Chen, and Rauscher (1995) and Goldin-Meadow and Sandhofer (1999) and for an alternative view.

These findings suggest that some properties of gestures may activate semantic information related to the content of the talk in progress. However, little is known about the cognitive and neural processes mediating this remarkable feat of multi-modal integration. Given growing interest in the role of motor mirroring systems in action comprehension (Rizzolatti & Arbib, 1998), the study of gesture may provide cognitive neuroscience with a further venue for understanding the relationship between sensori-motor and higher order conceptual processing.

It has been proposed by McNeill and others that during comprehension, speech and gesture are integrated into a common underlying conceptual representation. He writes,

* Corresponding author. Fax: +1 858 534 1128.
E-mail address: coulson@cogsci.ucsd.edu (S. Coulson).

0093-934X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2006.12.003

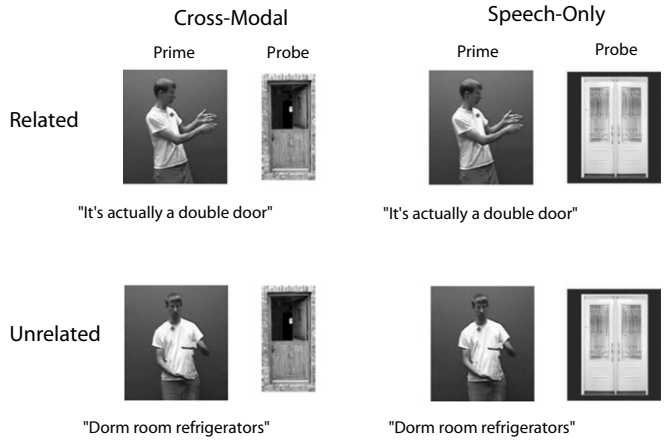

Fig. 1. Experimental design. Cross-modal and speech-only picture probes occurred in related and unrelated trials.

"…listeners, after a brief delay, cannot tell whether information came to them in gesture or in speech, the two having become unified (p. 17) (McNeill, 1998)." In support of this idea, a number of studies have investigated the comprehension of discourse in which the speakers' gestures express something different from their words (as in the Dutch door example). By assessing listeners' own accounts of what they had understood (Alibali, Flevares, & Goldin-Meadow, 1997; Cassell, McNeill, & McCullough, 1999; Goldin-Meadow, Wein, & Chang, 1992; Kelly & Church, 1998), or their responses on questionnaires (Goldin-Meadow & Sandhofer, 1999), it has been demonstrated that listeners are sensitive to information made available in both modalities.

The goal of the present study is to investigate how speech and gesture affect real-time interpretation processes. Previous behavioral research has demonstrated that when information is presented to listeners only through gesture, but not directly in speech, it is nevertheless accessible in long-term memory (for review, see Goldin-Meadow, 2003). However, little is known about the encoding processes whereby gesture-based information enters memory systems. Further, semantic activations induced by co-speech gestures have only begun to be investigated. The present study addresses the cognitive and neural processes mediating speech–gesture integration.

Recent research involving event-related potentials (ERPs) has begun to shed light on this question. ERPs represent dynamic voltage fluctuations that derive from synaptically generated current flow within patches of neural tissue. Tiny signals detectable at the scalp (on the order of microvolts) are amplified and digitized, yielding a record of on-going brain activity in the form of an electroencephalogram (EEG). By averaging portions of EEG recorded in synchrony with the presentation of a specific class of stimuli, it is possible to draw inferences about cognitive processes engaged by that type of stimulus. Because scalp-recorded potentials typically reflect contributions from a number of different neural sources, it is necessary to average event-related responses across many trials in order to cancel out random noise introduced by background neural activity. The resulting ERP waveform can be analyzed as a series of positive- and negative-going deflections (commonly referred to as components) which are characterized by their amplitude, time course and distribution across scalp electrode sites.

A component particularly relevant to semantic processing is the N400, which was discovered during early research on language processing (Kutas & Hillyard, 1980). Kutas and Hillyard recorded ERPs to the last word of sentences that either ended congruously (as in (1)), or incongruously (as in (2)).

(1) I take my coffee with cream and sugar.
(2) I take my coffee with cream and dog.

By averaging the signal elicited by congruous and incongruous sentence completions, respectively, these investigators were able to reveal systematic differences in the brain's electrical response to these stimulus categories occurring approximately 400 ms after stimulus onset. Subsequent research has shown that N400 components are generated whenever stimulus events induce semantic or conceptual processing. As such, many investigators have used the N400 component of the brain waves as a dependent variable in psycholinguistic experiments (for review, see Kutas, Federmeier, Coulson, King, & Muente, 2000).

To investigate the effect of gestures on language comprehension, Kelly, Kravitz, and Hopkins (2004) recorded ERPs elicited by spoken words articulated in synchrony with gestures that were either congruent and incongruent with word meanings. Stimuli were constructed by videotaping an actor as he gestured to either a tall, thin glass or a short, wide dish in front of him while saying one of four speech tokens—namely, *tall*, *thin*, *short* or *wide*. Gestures indicated the location of these two items, and also depicted either the height or width of their referent. In the matching condition, the actor's speech corresponded with both the object as well as the spatial dimension indicated in gesture. In the complementary condition, the speech token described a different dimension of the referent from that depicted by the gesture (e.g. *tall* uttered in accompaniment with a gesture indicating the thin diameter of the glass). In the mismatch condition, the speech token corresponded to one object while the gesture corresponded to the other. Finally, in the no gesture condition, speech was presented alone.

Results yielded early effects of gesture congruency (between 100 and 352 ms), with mismatching and complementary stimuli eliciting relative to other conditions larger P1 and P2 components, which reflect auditory sensory processing. N400-like effects were also observed at bilateral temporal electrode sites, with mismatch trials eliciting more negative ERPs than all other conditions around 450 ms post-stimulus. These findings suggest that gesture congruency affects both early sensory as well as higher order semantic processing of words.

Other studies have approached the neuro-cognitive underpinnings of gesture comprehension by measuring ERPs elicited by gestures themselves. Besides words, the

N400 component has also been elicited by image-based stimuli. For example, line drawings of objects elicit a lower amplitude N400 when they are preceded by a related word as compared with an unrelated one (Praterelli, 1994). Likewise, both line drawings and photographs of everyday items elicited less negative N400 when the preceding prime picture was a semantically related object relative to an unrelated one (Holcomb & McPherson, 1994; McPherson & Holcomb, 1999).

To test whether gestures engage semantic processes similar to those recruited by pictures and words, Gunter and Bach compared ERPs elicited by emblematic gestures (e.g. thumbs up, OK sign) and neutral hand postures that were of no conventional significance (Gunter & Bach, 2004). Previous research has shown that stimuli whose meaning cannot be accessed, as in the case of pseudowords (Holcomb, 1993) or unidentifiable objects (Holcomb & McPherson, 1994; McPherson & Holcomb, 1999), produce larger N400 than either semantically related or unrelated items. By analogy, neutral hand postures were found to elicit enhanced N400 relative to emblematic gestures. Similarly, spontaneous iconic gestures produced in the course of conversation elicited more negative ERPs between 400 and 600 ms post-stimulus (gesture N450) when preceded by incongruent contexts relative to congruent ones (Wu, 2005; Wu & Coulson, 2005). Due to its functional characterization, the gesture N450 was interpreted as a member of the N400 family of negativities. These findings suggest that in spite of their sometimes idiosyncratic, schematic, and dynamic qualities, gestures are subject to semantic processes.

Rather than examining the effect of gestures on speech processing, as in Kelly et al. (2004), or the effect of context on gesture processing, as in Wu and Coulson (2005), the goal of the present study was to investigate the respective contributions of speech and gesture to discourse comprehension. In keeping with McNeill (1998), we hypothesize that gestures activate semantic information related to the content of the talk in progress, allowing listeners to form a more robust representation of the speaker's intended meaning.

Research on the role of perceptual simulation in sentence comprehension offers experimental support for the view that language comprehension involves the activation of visuo-spatial properties even when they are not verbally expressed. Healthy adults have been presented, for example, with sentences followed by pictures showing an object configured in manner which was either congruent or incongruent with a given shape or orientation implied by within the sentence. An eagle with outstretched wings or folded ones followed *The ranger saw the eagle in the sky*. Similarly, either a horizontal or an upright pencil followed *John put the pencil in the drawer*. Pictures that were congruent with implied spatial information were named or classified more quickly than incongruent ones (Stanfield & Zwaan, 2001; Zwaan, Stanfield, & Yaxley, 2002). Studies such as these demonstrate that even in the absence of explicit encoding, visuo-spatial features including size, shape, and orientation are important for language comprehenders. When sufficient information is available from sentences to draw inferences about these features, they become active in the listeners' conceptual models. Iconic gestures, we propose, may also serve as an analogue resource for prompting such activations.

To test this proposal, we recorded ERPs as healthy adults viewed short segments of spontaneously produced discourse involving both descriptive speech and gestures. Our stimuli involved primarily iconic gestures because they typically express visuo-spatial relations that do not easily lend themselves to linguistic encoding (Emmorey & Casey, 2001). In order to assess semantic activations attributable to gestures, we measured the brain response to probe images that either did or did not reflect the spatial information conveyed by gestures in each discourse segment.

Besides our utilization of a real-time measure of neural processing, the present study differs from previous research on speech–gesture integration in at least two important ways. First, the discourse segments used in this research were derived from spontaneously produced conversation. Studies involving rehearsed speech–gesture mismatches created by actors (such as Cassell et al., 1999 or Kelly et al., 2004) may not be reflective of comprehension processes engaged during every day conversation. Secondly, the present study differs from much existing research using spontaneous gesture stimuli in that the speaker does not make reference to physically present objects or aspects of his immediate environment. This feature enabled us to test visuo-semantic activations prompted exclusively by gestures.

Each discourse prime was followed by one of four types of picture probes (see Fig. 2). Cross-modal related items were congruent with both the verbal and gestural component of the speaker's description. Speech-only related items were congruent only with descriptive features expressed through speech. A third logical "gesture-only related" condition—where pictures would be congruent with gestures, but not speech—was not included, though it would have afforded the opportunity to test for dissociable effects of gesture and speech relatedness on the N300 and N400. Instead, two types of unrelated stimuli were created by pairing cross-modal and speech-only related items with

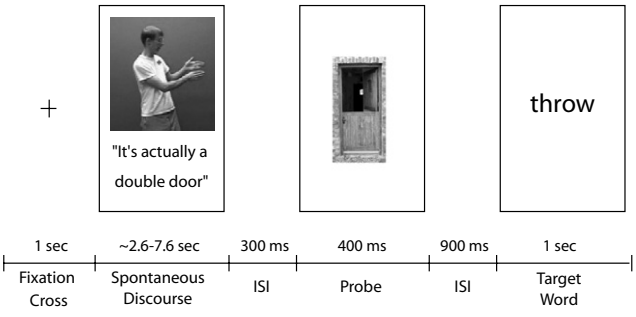

Fig. 2. Procedure. Discourse primes were followed by picture probes, and then target words.

other discourse primes, allowing each related picture to be compared with itself as an unrelated item.

We expected differences in the processing of cross-modal and speech-only probes to be reflected in two ERP-components sensitive to semantic relatedness between images and prior context. Previous studies comparing brain activity elicited by the second member of related and unrelated picture pairs have reported in addition to the N400, an earlier negative-going deflection of the ERP waveform peaking around 300 ms (N300) after stimulus onset. The amplitudes of the N300 and N400 are both larger (more negative) in response to unrelated items (Barrett & Rugg, 1990; Holcomb & McPherson, 1994; McPherson & Holcomb, 1999). This pattern of results has also been observed in experimental paradigms involving cross-modal priming between words and pictures (Hamm, Johnson, & Kirk, 2002), pictorial completions of written sentences (Federmeier & Kutas, 2001, 2002), and complex scenes (West & Holcomb, 2002). When viewed in concert, these findings suggest that the N300 and N400 index the integration of incoming visual semantic input with recently activated stored knowledge.

Although both the N300 and N400 are modulated by contextual congruity, they are thought to reflect slightly different aspects of image comprehension. Because the N300 has only been observed in response to photographs or pictures, it has been proposed to index image-specific semantic processes. By contrast, the N400 has been found in studies involving a broad range of meaningful representations, including words (see Coulson & Van Petten (2002) for review), ASL hand signs (Neville et al., 1997), gestures (Gunter & Bach, 2004; Wu, 2005; Wu & Coulson, 2005), and action videos (Sitnikova, Kuperberg, & Holcomb, 2003).

The N400 is also functionally distinct from the N300 in its sensitivity to degrees of semantic fit. Whereas the amplitude of the N300 reflects differentiation only between unrelated and related trials, the amplitude of the N400 increases in a graded fashion in response to highly related, moderately related, and unrelated items (McPherson & Holcomb, 1999). On the basis of these findings, it has been proposed that the N300 reflects image-specific semantic processes, while the N400 reflects more general semantic integration processes.

If the semantic activations engendered by speech and gesture jointly contribute to the ongoing formation of discourse level representations, then cross-modal pictures, which relate to information made available in both modalities, should be easier to process than speech-only pictures, which relate only to the speech, as indexed by the amplitude of the N300 and N400 ERP components. In general, we expect unrelated probes to elicit more negative ERPs than related probes. More importantly, we expect the size of relatedness effects to differ as a function of probe type. If pictures that are consistent with both speech and gestures are identified more readily than pictures consistent with speech alone, we would expect a larger cross-modal than speech-only relatedness effect within the time window of the N300. Further, if the semantic content of picture probes is easier to process when it agrees both with information expressed through speech and gesture relative to speech alone, we would expect a larger cross-modal than speech-only N400 related effect as well.

2. Methods

2.1. Participants

Sixteen volunteers (6 females and 10 males) were paid $16 or received academic course credit for their participation. All participants were healthy, fluent English speakers with no history of neurological impairment (mean age = 20, $SD = 2$). Fifteen individuals were right handed, and one was ambidextrous. The Edinburgh Inventory (Oldfield, 1971), revealed a mean laterality quotient of .73, in keeping with participants' self-reported right handed bias.

2.2. Materials

Stimuli included 168 video clips in which a speaker described a common object or event. In each case, the speaker's talk and gesture conveyed complementary, but not identical information. In one trial, for instance, he says, "Two throw pillows," while indicating in gesture that they are located at opposite ends of a couch. In another trial, he describes a hammer by saying, "the handle…the handle is wooden," while showing the object's horizontal orientation with his hands.

Video clips were followed by either related or unrelated picture probes. Related probes either agreed with both the speaker's speech and his gestures (cross-modal related), or they agreed with his speech alone (speech-only related). In the case of the throw pillows, the cross-modal related item was a sofa with pillows on either end, while in the speech-only related item, the probe depicted a sofa with adjacent throw pillows. In the case of the wooden hammer handle, the same hammer was shown at both a horizontal (cross-modal related) and vertical (speech-only related) orientation. Unrelated trials were constructed by pairing the same picture probes with different discourse primes, yielding a 2×2 factorial design with two levels of relatedness (related, unrelated) and two levels of stimulus type (cross-modal, speech-only), with each stimulus serving as its own control (see Fig. 1).

Video clips were constructed by filming a naive individual as he described everyday activities, as well as photographs of common objects and scenes to an off-camera interlocutor. Six recording sessions took place. He was told that the video footage would be utilized in the construction of stimuli for a subsequent memory experiment; no mention of gestures was made. Experimental materials involved instances in which the speaker's spontaneous gestures conveyed information over and above that in his speech. These instances were captured and digitized into short video clips ranging in length from 2.6 to 7.6 s. Gestures either

re-enacted elements of everyday actions (turning a knob, shaking out clothes, making the bed), or depicted affordances or spatial features of objects and scenes (the shape of a vase, the handles on a canvas bag, the location of a door). A total of 168 experimental clips were constructed, along with 7 filler and 2 practice clips.

Picture probes were constructed by collecting digital photographs from internet databases. Both picture probes and video frames were centered on a black background and subtended approximately 8.3° visual angle horizontally and 6.2° vertically. Within the videos, the speaker himself subtended approximately 5° to 6° vertically and 3° to 4.5° horizontally, and primarily the head, arms, and upper torso of the speaker were shown. Within the picture probes, depicted items subtended between 2.4° and 5° and were surrounded by a white background frame.

A normative study was conducted to evaluate the identifiability of probe materials. Twelve individuals viewed probes and either named the depicted objects or indicated that they could not identify them. On average, both probe types elicited names in 96% ($SD = 3.6\%$) of responses, indicating that the two probe types were well balanced in terms of identifiability.

To evaluate how consistently probes were named, we defined the *common name* for each picture as the word that occurred most consistently in our informants' responses. For example, when presented with the cross-modal probe depicting a Dutch door (see Fig. 2), three individuals responded with the word, "door." Other responses included "pub door," "sectional house door," and "wooden door to brick house." We selected *door* as the common name for this probe. After identifying the common name for each picture, we calculated the proportion of hits for that name out of the total number of responses received. On average, cross-modal pictures elicited their most common names in approximately 72% of responses ($SD = 25\%$), and speech-only pictures, in approximately 75% ($SD = 27\%$) of responses. This outcome suggests that the content of both types of picture probes was identified at a consistent rate well above chance. Further, on 60% of trials, the most common name elicited by cross-modal and speech-only probes was identical. Finally, we used Latent Semantic Analysis (Landauer & Dumais, 1997) to assess the degree of semantic similarity between the names elicited by each probe type. On a scale of 1 to −1 (with 1 indicating shared identity), the mean similarity between names was .69, ($SD = .4$).

To determine whether names elicited by each probe type were balanced for length and frequency, we used the Kucera-Francis database of written frequency counts to calculate the frequency of the most popular name for each trial. The mean frequency was 71 ($SD = 153$) for cross-modal names and 100 ($SD = 229$) for speech-only names. Differences in the variance within each name type appear to derive from the occurrence of a few very high frequency words in response to speech-only pictures. Differences in mean written frequencies were not statistically reliable ($t(335) = -1.4$, n.s.). The mean word length for cross-modal names was 5.8 ($SD = 2.3$) letters, and 5.6 ($SD = 2.3$) for speech-only names ($t(335) = .63$, n.s.).

A second normative study evaluated whether related and unrelated items could be reliably interpreted as such. Twenty additional volunteers listened to the digitized sound file extracted from each video clip and subjectively rated the degree of relatedness between each of the speaker's utterances and the subsequent picture probe on a scale of 1–5, with 5 designating the highest level. In the case of related trials, the mean rating was 4.3 ($SD = .7$) for cross-modal items, and 4.2 ($SD = .8$) for speech-only ones. In the case of unrelated trials, both sets of picture probes received mean ratings of 1.6 ($SD = .6$). As expected, related probes were rated as related to the prior context and unrelated probes were not. Moreover, when preceded by speech alone, cross-modal and speech-only related items were rated as equally related, and cross-modal and speech-only unrelated items were equally unrelated.

In order to encourage consistent attention to probe pictures, participants monitored for infrequent filler trials (7 total) in which the probes were dotted with blue paint splotches (applied through Adobe Photoshop). Data from these filler trials were not analyzed. Additionally, to provide an index of attention to discourse primes, each discourse-picture pair was followed by a single written task word which had either occurred in the immediately preceding speech, or was altogether new. Participants categorized each word as old or new by means of a button press, and response latencies and accuracy were measured, as in Wu and Coulson (2005).

Four randomized lists were constructed, each containing 42 cross-modal related items and 42 speech-only related items. Each list also contained 84 unrelated trials, wherein cross-modal and speech-only probes (42 each) were paired with unrelated video clips. No video or probe picture was repeated on any list, but across lists, each picture appeared once as a related stimulus, and once as an unrelated one. Equal numbers of new and old task words followed each type of related and unrelated trial.

2.3. Procedure

Each trial began with a fixation cross, presented in the center of a 17" color monitor for one second. Video clips were presented at a rate of 48 ms per frame and varied in total length (mean = 3752 ms, $SD = 1211$ ms). After a 300 ms pause, a picture probe appeared on the screen for 400 ms. Nine hundred ms after the offset of the probe, the written task word was presented for 1 s (see Fig. 2). A short pause (approximately 5 s) followed each trial as the next set of video frames was loaded for presentation.

Participants were told that they would watch a short video of a man describing something, followed by a picture, and then a word. They were asked to press YES on a button box if they had heard the word uttered previously, or else to press NO. Response hand was counterbalanced across subjects. They were also instructed to monitor for infrequent

blue splotches in picture probes, and were asked after each block if any had occurred.

2.4. EEG recording

The electroencephalogram (EEG) was recorded in a sound proof, electro-magnetically shielded chamber. Tin electrodes were used at 29 standard International 10–20 sites (Nuwer et al., 1999), including midline (FPz, Fz, FCz, Cz, CPz, Pz, Oz), medial (FP1, F3, FC3, C3, CP3 P3, O1, FP2, F4, FC4 C4, CP4, P4, O2), and lateral channels (F7, FT7, TP7, T5, F8, FT8, TP8, T6) (see Fig. 3). Electrodes were also placed on the right mastoid for off-line re-referencing, below the right eye for monitoring blinks, and a bipolar montage was placed at the outer canthi for monitoring horizontal eye movements. With the exception of the horizontal eye channels, all electrodes were referenced online to the left mastoid, and impedances maintained below 5 kΩ. EEG was amplified with an SA Instrumentation isolated bioelectric amplifier (band pass filtered, 0.01 to 40 Hz) and digitized on-line at 250 Hz. Data were later re-referenced to the algebraic mean of the left and right mastoids.

2.5. EEG analysis

Trials contaminated by artifacts such as blinks, eye-movements, blocking, and drift were rejected offline. Artifact-free trials were sorted and averaged, time-locked to the onset of picture probes. ERPs extended from 100 ms before stimulus onset to 920 ms after. On average, critical bins contained 35 trials (37 median). The mean artifact rejection rate was 14% ($SD = 12\%$).

Relatedness effects were assessed by measuring the mean amplitude (that is, the average of digitized voltage measurements obtained within a sampling window, calculated relative to the pre-stimulus baseline) and peak latencies (i.e., the time point when the amplitude reaches its maximal value) of ERPs for each subject. Time windows for measurement were 250–350 ms (N300) and 350–550 ms (N400) after stimulus onset—based on measurement intervals utilized in other studies involving picture probes (Federmeier & Kutas, 2002; McPherson & Holcomb, 1999), as well as visual inspection of the waveforms. Measurements were subjected to a 2×2 repeated measures ANOVA with the factors of Relatedness (probes were either related or unrelated to discourse primes), and Stimulus Type (cross-modal probes depicted information made available in speech and gesture; speech-only probes depicted information expressed in speech alone).

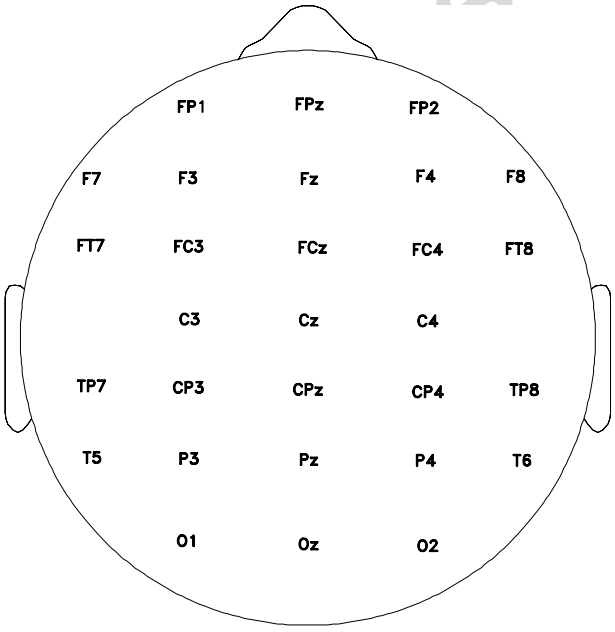

Fig. 3. Schematic diagram of scalp electrode sites.

To investigate the scalp distribution of ERP effects, an additional factor of Electrode Site (29 levels—corresponding to the 29 electrode channels) was included in the omnibus ANOVA. ERP effects qualified by an interaction with the electrode site factor were subject to three types off follow-up tests confined to specific groups of electrodes: Midline sites (with 7 levels along the anterior–posterior axis—namely, FPz, Fz, FCz, Cz, CPz, Pz, Oz), Medial sites (with 2 levels of hemisphere—left and right—and 7 anterior–posterior levels—FP1/FP2, F3/F4, FC3/FC4, C3/C4, CP3/CP4 P3/P4, O1/O2), and from Lateral sites (with 2 levels of hemisphere and 4 anterior–posterior levels—F7/F8, FT7/FT8, TP7/TP8, T5/T6). These follow-up analyses were designed to identify scalp regions where the effect was largest.

Additionally, we compared the topography of relatedness effects resulting from cross-modal and speech-only pictures by performing two point by point subtractions of ERPs elicited by related items from those elicited by unrelated items, yielding difference waves for each stimulus type. Repeated measures ANOVAs were performed on raw difference waves. For all analyses, original degrees of freedom are reported; however, where appropriate, p-values reflect Geisser–Greenhouse correction (Geisser & Greenhouse, 1959).

2.6. Control study

To rule out the possibility that ERP effects might derive purely from differences in visual complexity between cross-modal and speech-only stimuli, we recorded ERPs as six healthy adults (who did not participate in any other portion of this study) viewed picture probes presented in the absence of discourse primes. Two lists were constructed, each containing 84 cross-modal and 84 speech-only pictures. Pictures were followed by single written words (identical to those used in the main experiment). Participants were instructed to attend to all pictures, and to classify each word as either related or unrelated to the preceding picture by means of a button press.

Mean amplitudes of ERPs to each probe type were measured within the time windows used in the main experiment

to assess the N300 (250–350 ms post-stimulus) and N400 (350–550 ms post-stimulus). Because null effects were expected in this paradigm, ERPs elicited by related and unrelated probe words (subjectively categorized by participants) were also measured between 300 and 500 ms post-stimulus in order to ensure that the data set afforded sufficient power to detect statistically reliable differences. Picture and word ERP measurements were subjected to repeated measures ANOVA with the factors of Stimulus Type (cross-modal, speech-only) and Electrode Site for pictures, and relatedness (related, unrelated) and electrode site for words.

3. Results

3.1. Control study

In response to pictures, no main effects of stimulus type or interactions with electrode site were found in either the time window for the N300 (Main Effect: $F(1,5) = .03$, n.s.; Stimulus × Electrodes interaction: $F(28,140) = .8$, n.s.) or the N400 (Main Effect: $F(1,5) = .02$, n.s.; Stimulus × Electrodes interaction: $F(28,140) = .8$, n.s.). However, in response to probe words, unrelated words reliably elicited more negative ERPs than related ones (Relatedness Main Effect: $F(1,5) = 7.4$, $p < .05$). This outcome demonstrates sufficient power in our sample size to detect reliable differences in brain response. Although interpreting null results yielded by the picture probes should nevertheless be approached with caution, the absence of any effect of stimulus type on ERP amplitudes suggests that the visual properties of cross-modal and speech-only stimuli were well balanced.

3.2. Main experiment

3.2.1. Behavior

On average, participants accurately responded to 96% ($SD = .03$) of target words and 98% ($SD = .02$) of distractor words. A two-tailed t-test revealed that this small difference was nevertheless reliable ($t(15) = -2.5$, $p < .01$), suggesting a slight bias on the part of participants toward the *no* response. The mean response time for classifying targets was 927 ms ($SD = 322$), and 1002 ms ($SD = 316$) for distractors. This difference did not approach conventional significance, however ($t(30) = .67$, n.s.)—perhaps due to insufficient power. Overall, the high accuracy rates and trend towards an advantage for targets suggests that participants consistently attended to video primes.

3.2.2. ERPs to picture probes

Fig. 4 shows the effect of relatedness in cross-modal and speech-only conditions. For both stimulus types, a negativity peaking around 130 ms (N1) can be observed, followed by two subsequent negativities, which peak around 295 ms (N300) and 430 ms (N400), respectively. ERPs elicited by unrelated items diverge from their related counterparts after 250 ms in the cross-modal condition, and after 350 ms in the speech-only condition. For both types of stimuli, ERPs remain more negative in response to unrelated items relative to related ones in the latter portion of the epoch (after 550 ms).

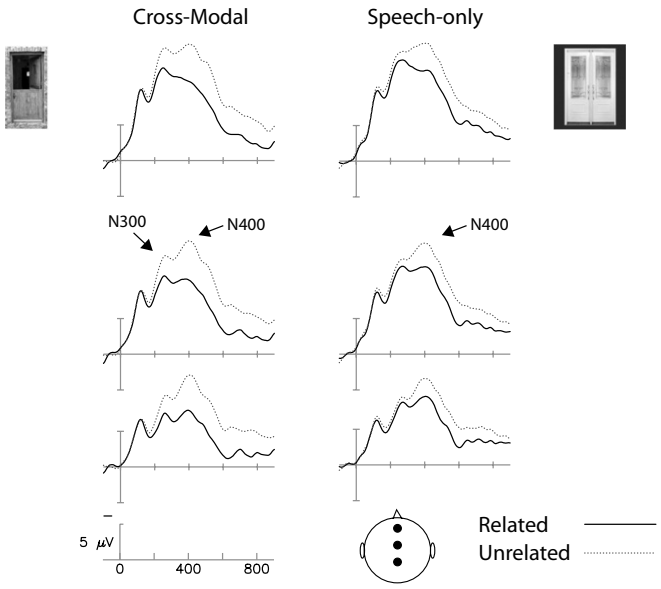

Fig. 4. ERPs recorded over central midline sites time-locked to the onset of picture probes. Negative polarity is plotted up, here and in all subsequent figures.

3.2.3. N300

Does N300 sensitivity to relatedness differ for cross-modal pictures as compared to speech-only ones? Overall, unrelated stimuli elicited more negative ERPs than related ones between 250 and 350 ms post-stimulus (Relatedness Main Effect: $F(1,15) = 28$, $p < .0001$), and speech-only stimuli elicited more negative ERPs than cross-modal ones (Stimulus Type Main Effect: $F(1,15) = 10$, $p < .01$). These main effects were qualified by a Relatedness × Stimulus Type interaction ($F(1,15) = 4.5$, $p = .05$), suggesting that the sensitivity of the N300 to probe relatedness differed as a function of whether probes were cross-modally consistent with both speech and gesture or consistent with speech alone.

What drives the interaction between relatedness and stimulus type? Follow-up analyses within each probe type revealed that cross-modal unrelated probes consistently elicited more negative ERPs than related ones (Relatedness Main Effect: $F(1,15) = 51$, $p < .0001$; Relatedness × Electrode Site: $F(28,420) = 5.1$, $p < .005$). For speech-only probes, by contrast, neither the main effect of relatedness nor the interaction with electrode site proved reliable (F's < 2, n.s.). These outcomes indicate that the visuo-semantic processes indexed by the N300 reliably distinguished between related and unrelated items only in the case of cross-modal stimuli.

Where was the N300 effect in response to cross-modal stimuli largest? The interaction with electrode site obtained in the simple contrast between cross-modal related and unrelated items indicated that their effect on

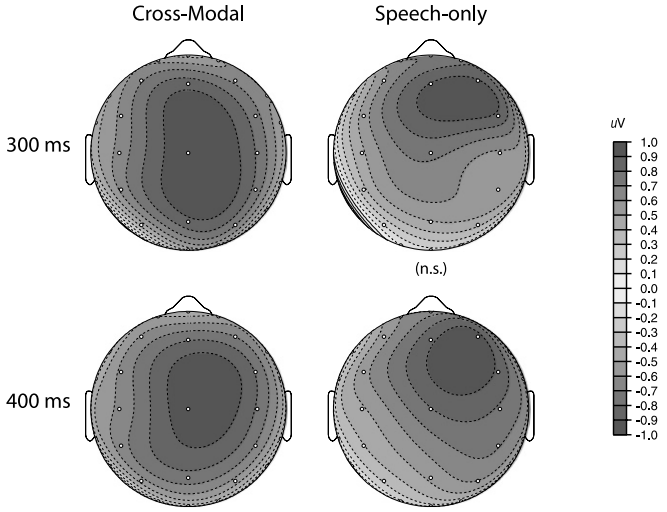

Fig. 5. Scalp topography of N300 and N400 relatedness effects (Unrelated minus Related) at 300 and 400 ms post-stimulus onset. The N300 effect was significant only in the cross-modal condition. For both cross-modal and speech-only conditions, values were normalized within a range of 1 and −1.

N300 amplitude was not uniform across the scalp. To characterize its distribution, follow-up analyses were conducted within midline, medial, and lateral electrode sites. In all three types of analyses, the N300 effect was most prominent over fronto-central electrode sites (Relatedness × Posteriority: Midline, $F(6,90)=3.6$, $p<.05$; Medial, $F(6,90)=3.7$, $p=.07$; Lateral, $F(3,45)=10.5$, $p<.005$) with a maximum over FCz and Cz. Further, the effect was larger over anterior right-hemisphere electrode sites than left-hemisphere ones (Relatedness × Hemisphere × Posteriority: Medial, $F(6,90)=2.9$, $p<.05$) (see Fig. 5 for scalp map of the N300 and N400 relatedness effects).

3.2.4. N400

Is the N400 also selectively sensitive to the semantic relatedness of cross-modal items? Between 350 and 550 ms, unrelated items consistently elicited more negative ERPs than related ones (Relatedness Main Effect: $F(1,15)=54.6$, $p<.0001$). The main effect was qualified by an interaction between relatedness and stimulus type ($F(1,15)=4.7$, $p<.05$). This result shows that N400 sensitivity to probe relatedness differed as a function of probe type.

Follow-up analyses within cross-modal and speech-only probes revealed that unrelated items elicited more negative ERPs than related ones in response to both types of stimuli (Cross-modal: Relatedness Main Effect, $F(1,15)=57.7$, $p<.0001$; Relatedness × Electrode Site, $F(28,420)=10.3$, $p<.0001$; Speech-only: Relatedness Main Effect, $F(1,15)=23.4$, $p<.0005$; Relatedness × Electrode Site, $F(28,420)=8.8$, $p<.0005$). However, the relatedness effect was larger in the cross-modal condition (3 μV) as compared to the speech-only condition (1.9 μV) (see Fig. 6). In other words, cross-modal related probes resulted in greater reduction of the N400 amplitude relative to unrelated controls than did speech-only probes.

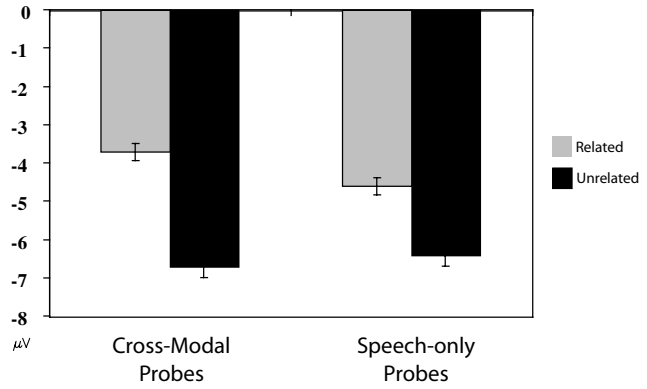

Fig. 6. The mean amplitude of N400 across all electrode sites (in microvolts) elicited by cross-modal and speech-only picture probes.

What was the topography of cross-modal and speech-only N400 effects? Given relatedness by electrode site interactions obtained in responses to both stimulus types, follow-up analyses were performed to further assess the distribution of relatedness effects. For cross-modal items, this effect was largest over anterior electrodes, as is typical of N400 elicited by pictorial stimuli (Ganis, Kutas, & Sereno, 1996; Holcomb & McPherson, 1994; West & Holcomb, 2002) (Relatedness × Posteriority: Midline, $F(6,90)=9.3$, $p<.001$; Medial, $F(6,90)=8.7$, $p<.005$; Lateral, $F(3,45)=9.6$, $p<.01$) with a central midline maximum, as in the previously measured window. For speech-only items, the relatedness effect was also frontally focused (Relatedness × Posteriority: Midline, $F(6,90)=7.7$, $p<.005$; Medial, $F(6,90)=6.4$, $p<.01$; Lateral, $F(3,45)=4.3$, $p=.05$). However, the effect was larger over the right than the left hemisphere (Medial: Relatedness × Hemisphere, $F(1,15)=6.3$, $p<.05$; Relatedness × Hemisphere × Posteriority, $F(6,90)=5$, $p<.01$; Lateral: Relatedness × Hemisphere, $F(1,15)=8.4$, $p<.05$) (see Fig. 5).

Do cross-modal and speech-only N400 effects reflect different distributions? Comparing difference waves of cross-modal and speech-only relatedness effects did not yield an interaction with electrode site during either the N300 ($F<1$) or N400 ($F<1$) time windows. This result indicates that the scalp distribution of relatedness effects elicited by each stimulus type did not reliably differ—despite subtle differences apparent in Fig. 5.

3.2.5. Control for discourse prime bias

Could differences between cross-modal and speech-only relatedness effects be driven purely by differences in the semantic fit between picture probes and the speech component of discourse primes? Although both cross-modal and speech-only stimuli were judged to be equally related to the speaker's utterances (see Section 2), it is possible that speech segments were more predictive of cross-modal pictures than speech-only ones, resulting in the observed reduction of N300 and N400 in response to cross-modal related items relative to speech-only ones. To rule out this possibility, we conducted a forced-choice normative study

in which 18 additional volunteers listened to the sound-file extracted from each video-clip and indicated whether the speaker's utterance corresponded better to the cross-modal or speech-only related picture. On the basis of participants' responses, ERP trials were sorted into two categories. Cross-modally biased trials were those in which the cross-modal picture was preferred over its speech-only counterparts by more than 55% of respondents. Unbiased trials were those which garnered fewer than 55% of responses in favor of the cross-modal probe.

ERPs elicited by biased and unbiased pictures were subjected to a $2 \times 2 \times 2$ repeated measures ANOVA with the factors of Bias, Relatedness, and Stimulus Type. Importantly, no main effect of Bias occurred within the time window of the N300 ($F < 1$, n.s.) or the N400 ($F < 1$, n.s.). Further, the factor of Bias did not interact with Relatedness or Stimulus Type within either time window (F's < 1, n.s.). These findings suggest that the degree of semantic fit between speech segments and picture probes was not responsible for the larger relatedness effects observed in response to cross-modal as compared to speech-only items.

4. Discussion

The brain response to cross-modal and speech-only probes differed in two ways. First, while cross-modal related items elicited less negative N300 than unrelated ones, no N300 effect was observed for speech-only probes. This outcome suggests that cross-modal related pictures were easier to identify than speech-only related ones. Second, although both related probe types elicited reduced N400 relative to unrelated controls, the N400 effect was larger to cross-modal probes. This result indicates that cross-modal related pictures fit the semantic context created by discourse primes more readily than speech-only related ones. Below we discuss the implications of the cross-modal N300 effect for perceptual priming by gestures. Further, we discuss our finding of the larger cross-modal N400 effect relative to McNeill's hypothesis that speech and gesture constitute an integrated system of thought.

4.1. N300 and image processing

In the present study, discourse primes served to modulate the N300 response to cross-modal but not to speech-only picture probes. These findings suggest co-speech gestures affect image-specific semantic processes indexed by the N300. However, differing views have been advanced regarding the functional significance of this ERP component. McPherson and Holcomb (1999) propose that it reflects the activation of image-specific semantic properties of objects. This idea is corroborated by the recent discovery that pictures of animals elicit larger N300 than pictures of tools over anterior electrode sites. Because identifying animals depends crucially on the interpretation of their visual characteristics, whereas tool identification relies more heavily on knowledge of an item's functional properties, the anterior N300 response to animals is thought to reflect heightened visuo-semantic processing prompted by animals relative to tools (Sitnikova, West, Kuperberg, & Holcomb, 2006).

Alternatively, given the finding that between-category violations (duck—collie) yield N300 effects, while within-category violations (poodle—collie) do not, the N300 has also been proposed to index a process whereby the structural properties of a percept are assigned to a generic basic-level category representation before more identity-specific information becomes available (Hamm et al., 2002). Moreover, Schendan and Kutas have reported an early anterior negativity similar to the N300 that is larger in response to unidentified relative to identified objects (Schendan & Kutas, 2002), as well as in response to objects presented from unusual as compared to canonical views (Schendan & Kutas, 2003). These findings have led to the proposal that the amplitude of this anterior negativity is modulated by the size of the search space of possible object representations to which the percept could be matched, with larger amplitudes elicited by images which could correspond to a wide range of possible interpretations.

Ultimately, any inferences about the N300 must take into consideration the semantic richness of the stimulus conditions under which it is observed. When expectations regarding the content of an upcoming image are highly constrained by preceding context, N300 sensitivity to within-category violations has been reported (Federmeier & Kutas, 2001, 2002), in contrast to the findings of Hamm et al. (2002). Similarly, in the present work, differential N300 response was elicited by pictures that were either consistent or inconsistent with visuo-spatial cues made available through gestures. Cross-modal related pictures resulted in reduced N300 relative to unrelated ones, while the N300 elicited by speech-only related and unrelated items did not reliably differ. In other words, during this time interval the brain treats all speech-only probes as being completely unrelated to their preceding context, but differentiates between cross-modal related and unrelated items.

If the N300 reflects cognitive processing mediating object recognition, the present findings suggest that iconic gestures served to aid the identification of cross-modal related stimuli. We propose that visuo-spatial cues provided by iconic gestures enabled listeners to formulate more precise conceptual representations of the items described in each utterance, thereby facilitating processes devoted to mapping percepts to stored knowledge and meaning for cross-modal related items, but not speech-only ones. This idea is consistent with current theories positing top-down facilitation of object recognition from low spatial frequencies in the image. It has been proposed that global shape information, such as orientation, size, and proportions, becomes available early during image processing—around ~100–250 ms (Schmid, Eddy, & Holcomb, 2005)—activating multiple possible high-level

representations that constrain the interpretation of bottom-up input (Bar, 2003).

By analogy, visuo-spatial features of the gestures in this experiment may also have pre-activated a range of high-level representations which made related cross-modal items easier to identify than their speech-only counterparts. This hypothesis could be further investigated by augmenting the present experiment with a gesture-only condition, whereby related picture probes would agree with the speaker's gestures, but not his verbal utterances. If gestures exert top-down influences on probe image recognition, then larger N300 effects should occur in the gesture-only than the speech-only conditions.

4.2. N400 and semantic integration

The larger N400 effect obtained in cross-modal trials relative to speech-only ones suggests that cross-modal related items were easier to interpret than their speech-only counterparts. Importantly, this processing difference can be attributed to the additional semantic cues supplied by gestures, as an off-line rating study revealed that both cross-modal and speech-only related items were rated as equally related to the speaker's utterances when his gestures were not shown. Additionally, when probe pictures were presented on their own, with neither speech nor gestures preceding them, both probe types elicited ERPs of similar amplitude during the N300 and N400 time windows. This finding discounts the possibility that ERP effects may have derived from differences in visual properties between cross-modal and speech-only stimuli.

The interaction between Stimulus Type and Relatedness in the present study was driven by a difference in the amplitude of the N400 elicited by the two types of related probes, while the unrelated probes elicited N400's of similar amplitude. This pattern suggests that the cross-modal N400 relatedness effect reflects facilitation of cross-modal related items rather than the detrimental impact of context on the processing of unrelated items. Inferences about these outcomes must be tempered, however, by the caveat that semantic integration processes indexed by the N400 may overlap temporally with object recognition processes indexed by the N300. Given this possibility, the greater magnitude of the cross-modal N400 effect relative to the speech-only one may be due at least in part to the differential magnitudes of the cross-modal and speech-only N300 effects. Again, an additional gesture-only condition in the present experimental paradigm would likely speak to this question. If the N300 and N400 reflect dissociable processes, images which are related only to the speaker's gestures should elicit less N300 than unrelated counterparts; however, no N400 relatedness would be expected.

Another issue deserving further exploration is the possibility that different relationships between gestures and speech used in the present experiment might affect listener comprehension in different ways. In some cases, gestures provided critical information denoting a certain kind of item within a class (e.g., a Dutch instead of a French door; a cupboard instead of a wall shelf; a stove knob instead of a door knob). In other cases, they portrayed salient visuo-spatial features of objects (e.g., the location of a logo on a T-shirt, the shape of vase, the degree of openness of a car window). Finally, some gestures demonstrated the manner of action execution (e.g., mixing with a spoon rather than an electric mixer, writing by hand rather than typing on a keyboard, painting with vertical rather than horizontal brush strokes).

It is possible that in response to cases where gestures provide substantive information beyond what is available through speech, listeners may formulate mental representations that are both visually and semantically more consistent with cross-modal probes relative to speech only ones. Consider an example in which the speaker demonstrates the shape of a tall, vertical cupboard while saying, "…and opposite that just kinda before the wall is another shelf." Here, any type of wall-mounted shelf would be congruent with the speaker's utterance alone, whereas a tall, upright cabinet or cupboard with shelves would be congruent with both his gesture and his speech.

On the other hand, in cases where the gesture simply elaborates information expressed through speech, it is possible that listeners activate representations that are visually more consistent with the cross-modal probe, but not semantically so. For example, in one trial the speaker says, "…a Nokia cell phone set at an angle," while indicating its orientation in gesture. In this instance, both the cross-modal and speech-only probes depicted the same cell phone, but at different orientations.

If the amount of additional information provided by gestures relative to speech affects speech–gesture integration, we might expect the two distinct types of discourse primes described above to differentially modulate ERP responses to picture probes. Because gestures that offer a great deal of additional information may result in more specific semantic activations (e.g., a tall cupboard with shelves instead of any type of shelves, or Dutch door instead of a French one), we might expect that when occurring in discourse primes, they may yield a larger N400 effect in the comparison between cross-modal and speech-only probes than in cases where the gesture served mainly to elaborate information expressed through speech. On the other hand, because both of these types of discourse primes allow the listener to formulate a more specific visual representation of the object or event being described than would be possible based on speech alone, we might expect comparably sized N300 effects for cross-modal versus speech-only probes in response to both prime types.

The present study confirms and extends existing experimental investigations of co-speech gesture integration. As noted above, researchers have previously reported that when speech and gesture convey different information, listeners are sensitive to both (Cassell et al., 1999; Goldin-Meadow et al., 1992). In prior work, listeners were shown to subsequently express gesturally conveyed meanings in

speech, and vice versa, suggesting that activations from both modalities engage a common underlying substrate. The current study corroborates this view by demonstrating that semantic activations induced by speech and iconic gestures jointly contribute to emerging conceptual representations constructed during discourse comprehension. The fact that participants' on-going brain response to visual stimuli was modulated to a greater degree by cross-modal than speech-only stimuli suggests that listeners made use of semantic relations expressed through gesture, even though this information was never made overt in speech. This finding supports McNeill's proposal that during comprehension, listeners integrate both linguistically and gesturally encoded meanings. Additionally, an important theoretical consequence of this work is the idea that gestures enable listeners to construct perceptually specific conceptual representations of the speaker's intended message.

This proposal parallels sentence processing research which demonstrates that individuals make use of incoming linguistic input in order to formulate precise expectations about upcoming words. It has been shown, for example, that in sentences which strongly favor a particular kind of lexical completion, definite and indefinite articles which agree grammatically (Wicha, Bates, Moreno, & Kutas, 2003) or phonologically (DeLong, Urbach, & Kutas, 2005) with the anticipated completion elicit reduced N400 in comparison with those which do not. These results have been construed as evidence for pre-activation of specific word representations before their actual presentation.

By analogy, we propose that the gestures in the current study may pre-activate representations of visuo-spatial features including orientation, location, shape, and size, as well as motoric features associated with specific patterns of action execution. It is possible that mechanisms such as conceptual integration (Fauconnier & Turner, 1998, 2002) or "mesh" (Glenberg, 1997; Glenberg & Robertson, 1999) mediate a process whereby perceptual or relational similarities between the gesture and the entity being described make available the relevant visuo-spatial representations.

5. Conclusion

This study used ERPs to measure semantic activations prompted by co-speech gestures. We found that segments of spontaneously produced discourse involving speech and gesture differentially primed picture probes that agreed with information conveyed either through both channels (cross-modal probes) or through speech alone (speech-only probes). Cross-modal probes elicited a larger N400 relatedness effect than did speech-only probes. Cross-modal probes also elicited an N300 relatedness effect, whereas speech-only ones did not. These findings support the proposal advanced by McNeill (1992) that listeners combine information from speech and gestures to arrive at an enhanced understanding of their interlocutor's meaning. They further suggest that iconic gestures activate image-specific information about the concepts which they denote.

Acknowledgments

Y. W. was supported by Institute for Neural Computation NIH Training Grant # MH 2002-07 and Center for Research on Language NSF Training Grant DC 000-16.

References

Alibali, M. W., Flevares, L., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: do teachers have the upper hand? *Journal of Educational Psychology, 89*, 183–193.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience, 15*(4), 600–609.

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

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

Beattie, G., & Shovelton, H. (2002). An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *British Journal of Psychology, 93*(2), 179–192.

Cassell, J., McNeill, D., & McCullough, K. (1999). Speech-gesture mismatches: evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics and Cognition, 7*(1), 1–33.

Coulson, S., & Van Petten, C. K. (2002). Conceptual integration and metaphor: an event-related potential study. *Memory & Cognition, 30*, 958–968.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilitstic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 117–1121.

Emmorey, K., & Casey, S. (2001). Gesture, thought, and spatial language. *Gesture, 1*(1), 35–50.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive Science, 22*, 133–187.

Fauconnier, G., & Turner, M. (2002). *The way we think*. New York: Basic Books.

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 202–224.

Federmeier, K. D., & Kutas, M. (2002). Picture the difference: eletrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia, 40*, 730–747.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": an electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience, 8*(2), 89–106.

Geisser, S., & Greenhouse, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95–112.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20*, 1–55.

Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes, 28*(1), 1–26.

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

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

Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge through gestures: using children's hands to read their minds. *Cognition and Instruction, 9*(3), 201–219.

Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology, 10*, 57–67.

Gunter, T. C., & Bach, P. (2004). Communicating hands: ERPs elicited by meaningful symbolic hand postures. *Neuroscience Letters, 372*, 52–56.

Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to pictue stimuli in congruent and incongruent contexts. *Clinical Neurophysiology, 113*, 1339–1450.

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

Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition, 24*, 259–276.

Kelly, S. D. (2001). Broadening the units of analysis in communication: speech and nonverbal behaviours in pragmatic comprehension. *Journal of Child Language, 28*, 325–349.

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., & Church, R. B. (1998). A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child Development, 69*(1), 85–93.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*(1), 243–260.

Krauss, R. M., Dushay, R. A., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology, 31*, 533–552.

Kutas, M., Federmeier, K., Coulson, S., King, J. W., & Muente, T. F. (2000). Language. In G. G. Berntson (Ed.), *Handbook of psychophysiology* (2nd ed., pp. 576–601). Cambridge, UK: Cambridge University Press.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207*, 203–205.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211–240.

McNeill, D. (1992). *Hand and mind*. Chicago: Chicago University Press.

McNeill, D. (1998). Speech and gesture integration. In J. Iverson & S. Goldin-Meadow (Eds.), *The nature and functions of gesture in children's communication. New directions for child development* (Vol. 79, pp. 11–27). San Francisco: Jossey-Bass.

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology, 36*, 53–65.

Morford, M., & Goldin-Meadow, S. (1992). Comprehension and production of gesture in combination with speech in one-word speakers. *Journal of Child Language, 19*(3), 559–580.

Neville, H. J., Coffey, S. A., Lawson, D. S., Fischer, A., Emmorey, K., & Bellugi, U. (1997). Neural systems mediating American Sign Language: effects of sensory experience and age of acquisition. *Brain and Language, 57*(3), 285–308.

Nuwer, M. R., Comi, G., Emerson, R., Fuglsang-Frederiksen, A., Guerit, J. M., Hinrichs, H., et al. (1999). IFCN standards for digital recording of clinical EEG. The international federation of clinical neurophysiology. *Electroencephalography and Clinical Neurophysiology Supplement, 52*, 11–14.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh Inventory. *Neuropsychologia, 9*, 97–113.

Praterelli, M. E. (1994). Semantic processing of pictures and spoken words: evidence from event-related brain potentials. *Brain and Cognition, 24*, 137–157.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neuroscience, 21*, 188–194.

Rogers, W. T. (1978). The contribution of kinesic illustrators toward the comprehension of verbal behavior within utterances. *Human Communication Research, 5*, 54–62.

Schendan, H. E., & Kutas, M. (2002). Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia, 40*, 931–945.

Schendan, H. E., & Kutas, M. (2003). Time course of processes and representations supporting visual object identification and memory. *Journal of Cognitive Neuroscience, 15*(1), 111–135.

Schmid, A., Eddy, M., & Holcomb, P. J. (2005). Integration of bottom-up and top-down processes in visual object recognition. Paper presented at the Cognitive Neuroscience Society: Annual Meeting Program 2005, New York, NY.

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science, 16*(2), 85–89.

Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: an electrophysiological investigation. *Psychophysiology, 40*(1), 160–164.

Sitnikova, T., West, W. C., Kuperberg, G., & Holcomb, P. J. (2006). The neural organization of semantic memory: electrophysiological activity suggests feature-based segregation. *Biological Psychology, 71*, 326–340.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12*(2), 153–156.

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

Thompson, L. A., & Massaro, D. W. (1994). Children's integration of speech and pointing gestures in comprehension. *Journal of Experimental Child Psychology, 57*, 327–354.

Valenzeno, L., Alibali, M. W., & Klatzky, R. (2003). Teacher's gestures facilitate students' learning: a lesson in symmetry. *Contemporary Educational Psychology, 29*, 187–204.

West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research, 13*(3), 363–375.

Wicha, N. Y. Y., Bates, E., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters, 346*, 165–168.

Wu, Y.C. (2005). Meaning in gestures: what event-related potentials reveal about processes underlying the comprehension of iconic gestures. Retrieved August, 2005, from <http://www.crl.ucsd.edu/newsletter/>.

Wu, Y. C., & Coulson, S. (2005). Is that a meaningful gesture: electrophysiological indices of gesture comprehension. *Psychophysiology, 42*(6), 654–667.

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168–171.