The Role of Iconic Gestures in Speech Disambiguation: ERP Evidence

Henning Holle and Thomas C. Gunter

Abstract

■ The present series of experiments explored the extent to which iconic gestures convey information not found in speech. Electroencephalogram (EEG) was recorded as participants watched videos of a person gesturing and speaking simultaneously. The experimental sentences contained an unbalanced homonym in the initial part of the sentence (e.g., *She controlled the ball* . . .) and were disambiguated at a target word in the subsequent clause (*which during the game* . . . vs. *which during the dance* . . .). Coincident with the initial part of the sentence, the speaker produced an iconic gesture which supported either the dominant or the subordinate meaning. Event-related potentials were time-locked to the onset of the target word. In Experiment 1, participants were explicitly asked to judge the congruency between the initial homonym–gesture combination and the subsequent target word. The N400 at target words was found to be smaller after a congruent gesture and larger after an incongruent gesture, suggesting that listeners can use gestural information to disambiguate speech. Experiment 2 replicated the results using a less explicit task, indicating that the disambiguating effect of gesture is somewhat task-independent. Unrelated grooming movements were added to the paradigm in Experiment 3. The N400 at subordinate targets was found to be smaller after subordinate gestures and larger after dominant gestures as well as grooming, indicating that an iconic gesture can facilitate the processing of a lesser frequent word meaning. The N400 at dominant targets no longer varied as a function of the preceding gesture in Experiment 3, suggesting that the addition of meaningless movements weakened the impact of gesture. Thus, the integration of gesture and speech in comprehension does not appear to be an obligatory process but is modulated by situational factors such as the amount of observed meaningful hand movements. ■

INTRODUCTION

While engaged in face-to-face communication, speakers produce meaningful hand movements (i.e., gestures) as they speak, resulting in a visual and an auditory stream of information. We were interested in investigating how a listener incorporates information from each channel. Does the listener combine both streams of information in comprehension? If so, what factors determine the degree to which gesture is taken into account?

Co-speech gestures differ in their relationship to speech. Some gestures, called emblems (e.g., the o.k. sign), are so conventionalized in their form that they can be understood independently of a speech context (Gunter & Bach, 2004). Other gestures, such as pointing, are highly dependent on some kind of (speech) context to acquire meaning because the form of a pointing movement has no meaning per se. Iconic gestures take an intermediate position in that they are *somewhat* dependent on the speech context. For example, a speaker might produce typing hand movements accompanying a sentence like, ''He has finally started writing.'' In this case, the gesture illustrates a property of its corresponding speech unit ''writing.'' On the one hand, an iconic gesture is, in some ways, dependent on the speech context as gesture and speech both relate to the same semantic concept ''writing.'' Also, there is a temporal dependency between the iconic gesture and the corresponding speech unit: Speakers have a high tendency to produce the stroke of a gesture at the onset of the corresponding speech unit (McNeill, 2005; Nobe, 2000; Levelt, Richardson, & la Heij, 1985). On the other hand, iconic gestures are also independent of speech because the form of an iconic gesture has some meaning per se. In the example provided, the gesture indicates that writing was performed on a keyboard and not by pen and paper. This additional information can only be found in the iconic gesture. It has been a matter of debate in recent years whether listeners are able to benefit from this kind of additional gesture information in comprehension.

One suggestion has been that gestures are generated as an epiphenomenon of speech production processes, but have little value for the listener (Krauss, Dushay, Chen, & Rauscher, 1995). According to this stance, gestures primarily facilitate the speaker's lexical access but are only subject to minimal semantic analysis in

Max Planck Institute for Human Cognitive and Brain Sciences

comprehension. There are some data supporting this view. For example, participants have difficulties in selecting the correct corresponding speech unit of an iconic gesture (Hadar & Pinchas-Zamir, 2004; Krauss, Morrel-Samuels, & Colasante, 1991; Feyereisen, Van de Wiele, & Dubois, 1988). In another study, it was found that listeners did not benefit from the additional gesture information (Krauss et al., 1995).

An opposing view put forward by McNeill (1992) holds that iconic gestures do convey additional information to the listener. In McNeill's model, it is assumed that gesture and speech are part of a tightly integrated system. Gesture and speech each convey some unique and some redundant information, and the comprehension system routinely combines the bimodal information into an enriched unified representation. Thus, the model assumes an obligatory interaction between gesture and speech in the comprehension process. It predicts that iconic gestures are easily decoded and have a high impact on speech comprehension.

Evidence suggesting such a strong impact of gesture has mainly been obtained in two different experimental paradigms. In one approach, the effect of bimodal presentation (speech with accompanying gestures) is contrasted with the effect of unimodal presentations (speech only; gesture only). The dependent variable in these experiments is typically some measure of comprehension (e.g., the amount of recalled details). Participants can recall more events after bimodal presentation (Beattie & Shovelton, 1999a, 1999b, 2001, 2002), especially if gestures give some information about the relative size or position of objects (Beattie & Shovelton, 1999b). The bimodal-versus-unimodal paradigm allows, however, only limited inferences about the processes underlying gesture–speech integration. For instance, one cannot rule out the possibility that the advantage of the bimodal presentation is partly due to more attentive speech processing. Another drawback of the paradigm is the poor temporal resolution, which makes it impossible to determine whether gesture–speech integration is achieved by fast on-line or slower off-line processes.

Further data in support of a strong impact of gesture come from another approach, in which gesture and speech provide clearly conflicting information. In these so-called mismatch paradigms, the dependent variable is the amount of interference caused by the incompatible gesture–speech combination. An interference effect is taken as evidence that the comprehension system attempted to integrate gesture and speech. A great advantage of the mismatch paradigm is that it allows for the investigation of the time course of gesture–speech integration, for example, by analyzing event-related potentials (ERPs) from the electroencephalogram (EEG). All of the previously conducted ERP studies on gesture–speech integration focused on the N400 component (Özyürek, Willems, Kita, & Hagoort, 2007; Wu & Coulson, 2005; Kelly, Kravitz, & Hopkins, 2004). The N400, a negative-going waveform with its peak amplitude around 400 msec after stimulus onset, is associated with semantic processing. The more easily a word can be integrated into a sentence, the smaller the amplitude of the N400 (for a review, see Hinojosa, Martin-Loeches, & Rubia, 2001). In the study by Kelly et al. (2004), participants saw video clips of a person that gestured to one of two objects in front of him, namely, a short, wide dish and a tall, thin glass. Directly after the offset of the gesture, one of four speech tokens was auditorily presented, namely, *tall*, *thin*, *short*, or *wide*. ERPs were time-locked to the onset of the speech tokens. The N400 was found to be smaller when gesture and speech referred to the same object and larger when they referred to different objects (e.g., gesturing *tall* and saying *wide*). This result suggests that it is difficult to integrate an incongruent target word into a gesture context. In an experiment by Wu and Coulson (2005), participants judged the relatedness between probe words and preceding cartoon–gesture pairs. In the ERPs time-locked to the gestures, they reported an enhanced negativity in the N400 time range for incongruent gestures, suggesting that it is more difficult to integrate an incongruent gesture into a cartoon context. The study by Özyürek et al. (2007) directly compared the time course of semantic integration for gesture and speech. In this experiment, participants were presented with an initial sentence context that was subsequently matched or mismatched either in the gesture channel, in the speech channel, or in both. The synchrony between gesture and speech was manipulated so that the stroke onset of gesture always coincided with the onset of the target word. All mismatch conditions showed similar N400 effects. The authors concluded that the time window of semantic integration is similar for gestures and speech. In sum, all three ERP studies have provided important evidence that iconic gestures have an impact on on-line brain measures that are associated with semantic processing. However, studies employing a mismatch paradigm are somewhat limited in their external validity because speakers do not produce such clearcut gesture–speech mismatches in spontaneous conversations. Another disadvantage of the mismatch paradigm is that it focuses on potential conflict between gesture and speech in comprehension but does not address the issue of how gesture may aid language comprehension.

Taken together, the discussed data favor a communicative view of gesture in comprehension. Listeners seem to be able to retrieve some additional information from co-speech gestures. However, one open question is through which mechanisms gesture facilitates language comprehension. One possible mechanism may be that listeners use gestural information to disambiguate speech.

Indeterminacy is one of the most significant challenges in language comprehension. We are constantly

required to disambiguate stimuli with uncertain identities using whatever environmental and experiential context is available (see also Twilley & Dixon, 2000). Despite this massive ambiguity, selecting the contextually appropriate interpretation is an effortless process for a listener suggesting that our comprehension system is very efficient in disambiguation. One frequent kind of ambiguity is lexical ambiguity. For example, a sentence such as *The woman observed the ball* is lexically ambiguous because the contained homonym allows two plausible interpretations. A study by Holler and Beattie (2003) investigated the role of gesture in disambiguation. In this experiment, participants were asked to read sentences containing an underlined homonym. After each sentence, the experimenter asked which of the two word meanings the sentence referred to. In almost half of all explanations, participants produced co-speech gestures to illustrate the relevant meaning. Thus, gestures produced in the context of a homonym are a phenomenon that actually occurs in face-to-face conversations. The question of the current study is whether listeners make use of this gestural information in comprehension.

Whereas some homonyms have equally frequent meanings, most homonyms are unbalanced (e.g., *ball*) in that they have a more frequent dominant meaning (e.g., *game*) and a lesser frequent subordinate meaning (e.g., *dance*). During the processing of unbalanced homonyms, the comprehension system can use two sources of information to activate the appropriate word meaning: (1) the context in which the homonym is encountered and (2) word meaning frequency. The vast literature on this topic (for a review, see Twilley & Dixon, 2000) allows some predictions about how these two sources of information interact in activating the word meanings of a homonym.[1]

In the absence of context or in a neutral context, word meaning frequency determines the activation pattern. In such a situation, the dominant meaning is activated to a stronger degree than the subordinate meaning (Vu, Kellas, & Paul, 1998; Simpson & Krueger, 1991; Simpson & Burgess, 1985; Simpson, 1981). For example, in the experiment by Simpson and Burgess (1985), homonyms were presented without prior context. Participants made lexical decisions to target words that were associates of the dominant or the subordinate meaning of a homonym prime. The pattern of reaction times for the different SOAs indicates that the dominant meaning was activated more quickly and maintained longer than the subordinate meaning. Even when participants were explicitly asked to think of all possible meanings of a homonym, the dominant meaning was found to be more active than the subordinate meaning, suggesting that the activation process is not under strategic control (Simpson & Burgess, 1985, Experiment 3). Thus, in the absence of a contextual cue, word meaning frequency determines how the different meanings of a homonym are activated.

Once the homonym is preceded by a biasing sentence context, the activation of the subordinate word meaning always seems to be affected. It has consistently been reported that the subordinate meaning is more active after a congruent subordinate context and less active after an incongruent dominant context (Vu et al., 1998; Paul, Kellas, Martin, & Clark, 1992; Simpson & Krueger, 1991; Onifer & Swinney, 1981). Thus, the activation of the subordinate word meaning of a homonym varies reliably as a function of context congruency.

In contrast, the activation of the dominant word meaning does not always seem to be affected by context congruency. Whereas some studies reported that the dominant meaning was more active after a dominant context and less active after a subordinate context (Vu et al., 1998; Paul et al., 1992; Simpson & Krueger, 1991; Onifer & Swinney, 1981), others have found that the dominant meaning is always active, after both a dominant context as well as after a subordinate context (Tabossi, 1988; Tabossi, Colombo, & Job, 1987). These seemingly contradictory findings can be nicely explained by data from Martin, Vu, Kellas, and Metcalf (1999) and Simpson (1981). These two studies systematically varied the degree to which a preceding sentence context biased either the dominant or the subordinate meaning of a homonym. Their results show that only a strongly biasing context was able to modulate the activation of the dominant meaning, that is, the dominant meaning was more active after a strong dominant context and less active after strong subordinate context. Both of the weakly biasing contexts activated the dominant meaning to a similar degree, that is, the weak subordinate context was as effective in activating the dominant meaning as the weak dominant context. Thus, it seems to be the case that once the contextual constraints become weak, the comprehension system makes increased use of other sources of information, in this case, word meaning frequency. As a result, the dominant word meaning is always activated after weak contexts, even if the context biased the subordinate meaning.

Taken together, the literature on homonym processing suggests that a biasing context generally modulates the activation of the subordinate word meaning. Modulatory context effects of the dominant word meaning are restricted to strongly biasing contexts. It is a characteristic feature of weakly biasing contexts that they are unable to modulate the activation of the dominant word meaning. Instead of using simple sentences, the present study investigates the extent to which co-speech iconic gestures can constitute a contextual cue for homonym disambiguation. The use of unbalanced homonyms has, in this case, the particular advantage that one can infer from the observed pattern of results whether gesture had a strong or a weak impact on disambiguation. If gestures have the status of a strong contextual cue for a listener, the activation of both the dominant and the subordinate word meaning should vary as a function of

the congruency of the preceding gesture context (*strong context pattern*). If, however, iconic gestures have only a weak impact on disambiguation, only the activation of the subordinate word meaning should be modulated by context congruency (*weak context pattern*).

Another open question in co-speech gesture comprehension is the degree to which the integration between both modalities is an obligatory or ''automatic'' process. Does the listener always take gesture into account as has been suggested by McNeill, Cassell, and McCullough (1994)? Or are there situations in which the information from the gesture channel has no detectable influence on comprehension? In terms of experimental manipulations, there are different ways of addressing this issue. One way is to explore whether the effect of gesture is task-independent. For example, both the studies by Özyürek et al. (2007) and Kelly et al. (2004) found an interference effect, although the task in these experiments did not force the participants to take gesture into account. These findings suggest that at least in situations where gesture and speech provide clearly conflicting information, the interaction between both domains is obligatory.

A recent study by Kelly, Ward, Creigh, and Bartolotti (in press) approached the question of automaticity in a different way. The experimental paradigm was similar to the previously described study by the same first author (Kelly et al., 2004), but contained an additional manipulation of the intentional relationship between gesture and speech. Gesture and the subsequent target word were either produced by the same speaker or by two different speakers. When participants knew that the same speaker produced gesture and speech, the processing of mismatching versus matching words elicited a bilateral N400 effect. In contrast, when participants knew that gesture and the target word were produced by two different speakers, the N400 effect had a markedly different topography and was significant only at right frontal electrode sites. This can be seen as initial evidence that the processing of gesture and speech may not be an entirely automatic process.

A third possibility of testing the relative amount of automaticity is to manipulate the proportion of related versus unrelated prime–target combinations in an experiment. According to two-process theories of information processing (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977; Posner & Snyder, 1975), automatic processes are fast-acting, occur without intention or awareness, and do not use limited-capacity resources. In contrast, controlled processes are slower, are under a person's strategic control, and use limited-capacity resources. On the basis of these assumptions, several studies that sought the automaticity of semantic priming have used a relatedness proportion manipulation (Chwilla, Brown, & Hagoort, 1995; Koyama, Nageishi, & Shimokochi, 1992; Holcomb, 1988). In these experiments, it was tested whether a given indicator of priming (e.g., N400 effect for unrelated vs. related targets) varied as a function of the proportion of related word pairs within an experiment. If the N400 effect was unaffected by the relatedness manipulation, it was concluded that semantic priming is a primarily automatic process. If, however, the N400 effect was smaller in the context of many unrelated word pairs, it was suggested that semantic priming involves a considerable degree of controlled processes. Hahne and Friederici (1999) adopted a similar strategy in determining the automaticity of two different syntactic ERP components.

One possibility to apply a relatedness proportion manipulation to the field of gesture comprehension is the use of hand movements that are unrelated to the contents of speech. One obvious candidate for these meaningless hand movements are the self-touching movements speakers frequently produce (e.g., grooming). Based on the logic outlined above, the addition of meaningless grooming movements to an experiment on co-speech gesture comprehension constitutes a test for the automaticity of gesture–speech integration. If gesture–speech integration is a primarily automatic process, the addition of grooming movements should have no impact on the effect of gesture. If, however, controlled (e.g., strategical) processes are also substantially involved in the integration of gesture and speech, the addition of meaningless hand movements should weaken the impact of gesture in comprehension.

The Present Study

The present study examines whether co-speech iconic gestures influence the disambiguation of homonyms in auditory sentence processing. To this end, EEG was recorded as participants watched videos of a person simultaneously gesturing and speaking. The experimental sentences contained an unbalanced homonym in the initial part of the sentence (e.g., *Sie beherrschte den Ball ... / She controlled the ball ...*) and were disambiguated at a target word in the subsequent clause (*was sich im Spiel ... / which during the game ...* vs. *was sich im Tanz ... / which during the dance ...*). Coincident with the initial part of the sentence, the speaker produced an iconic gesture, which supported either the dominant or the subordinate meaning. ERPs were time-locked to the onset of the target word.

The literature cited above suggests a systematic relation between the observed activation pattern and context strength. If iconic gestures have the status of a strong contextual cue for a listener, the N400 time-locked to both dominant and subordinate target words should vary reliably as a function of context congruency. More precisely, the N400 time-locked to the subordinate target words should be smaller after a congruent subordinate gesture context and larger after an incongruent dominant gesture context. Conversely, the N400 time-locked to the dominant target words should be smaller

after a congruent dominant gesture context and larger after an incongruent subordinate context. However, if iconic gestures constitute a weak contextual cue, only the N400 time-locked to the subordinate target words should vary as a function of context congruency. Based on the literature, we hypothesize that iconic gestures have a strong impact on speech disambiguation.

Experiment 1 investigates whether iconic gestures are used as disambiguation cues using a task that forces participants to combine gesture and speech. Experiment 2 explores whether the findings of Experiment 1 are replicable once the task no longer forces participants to combine gesture and speech. Finally, Experiment 3 addresses the extent to which the disambiguation of homonyms through gesture is an automatic process by adding meaningless hand movements to the paradigm.

EXPERIMENT 1

Methods

Participants

Twenty-seven native German-speaking students were paid €7.5 per hour for their efforts, and signed a written informed consent. Three participants had to be excluded based on rejection criteria. The remaining 24 participants (14 women, mean age = 25 years, range = 21–30 years) were right-handed (mean laterality coefficient = 93; Oldfield, 1971). All participants had normal or corrected-to-normal vision, and none reported any known hearing deficit.

Stimuli

Homonyms. The present study is based on a set of 91 unbalanced German homonyms (for a description on how the set was obtained, see Gunter, Wagner, & Friederici, 2003). Each of the homonyms had a more frequent dominant and a lesser frequent subordinate meaning, which shared identical phonological and orthographical surface features (e.g., *ball*—dominant meaning: *game*; subordinate meaning: *dance*). Target words representing the dominant meaning as well as target words representing the subordinate meaning were assigned to each of the homonyms. The relatedness of the target words to the homonyms had been previously tested using a lexical decision task in the visual modality (see also Wagner, 2003). For all target words, the lexical decision time was significantly shorter as compared to an unrelated item. In cases where either the dominant or the subordinate meaning was very abstract, the homonym was excluded from the set, resulting in a reduced set of 55 homonyms. For each of these 55 homonyms, two 2-sentence utterances were constructed including either the dominant or the subordinate target word. The utterances consisted of a short introductory sentence introducing a character followed by a longer complex sentence describing an action of that character. The complex sentence was composed of a main clause containing the homonym and a successive subclause containing the target word. Previous to the target word, the sentences for the dominant and subordinate versions were identical (see Table 1).

Gesture recording. A professional actress was videotaped while uttering the sentences. The exact recording scenario was as follows. The actress stood in front of a video camera with her hands hanging comfortably in a resting position. In a first step, she memorized one 2-sentence utterance until she could utter it fluently. Then she was asked to utter the sentence and simultaneously perform a gesture that supported the meaning of the sentence. The gestures were created by the actress and not choreographed in advance by the experimenter. She was instructed to perform the gesture to coincide with the initial part of the complex sentence (e.g., *Sie kontrollierte den Ball/She controlled the ball*) and to return her hands to the resting position afterwards. About two thirds of all gestures reenacted the actions in the sentence from a first-person perspective (typing on a keyboard, swatting a fly, peeling an apple) while the remainder of gestures typically depicted salient features of objects (the shape of a skirt, the height of a stack of letters). To minimize influences of mimic, the face of the actress was covered with a nylon stocking. All gestures resembling emblems or gestures directly related to the target words were excluded.

Pretest. The selected video material was edited using commercial editing software (Final Cut Pro 5). A pretest was conducted to assess how effective the gestures were in disambiguating the homonyms. In this modified cloze procedure, the videos were displayed to 20 German native speakers with sound muted one word before the onset of the target word. The participants had to select the most probable sentence continuation. The two alternatives on the response sheet were the dominant and the subordinate sentence continuations (e.g., *was sich im Spiel/which during the game* vs. *was sich im Tanz/which during the dance*). Overall, the gestures elicited a cloze probability of 93.7%, which was significantly above chance level ($p < .01$). Only homonyms which could be disambiguated by gesture in at least 80% of all participants were kept, resulting in the final set of 48 homonyms. In this final set, dominant and subordinate gestures did not differ significantly in their cloze probability [paired $t(1, 47) = 0.69, p > .4$].

Splicing. The speech of the sentences was re-recorded in a separate session to improve the sound quality. Because listeners may also use prosodic cues to resolve lexical ambiguities, half of the sentences were realized

Table 1. Stimulus Examples

Gesture	*Target Word*	*Gesture/Homonym*	*Target Word*
Introduction: Alle waren von Sandra beeindruckt.			
Everybody was impressed by Sandra.			
D	D	Sie kontrollierte den $Ball_{amb}$, was sich im	**Spiel** beim Aufschlag deutlich zeigte.
		She controlled the $ball_{amb}$, which during the	***game*** *at the serve clearly showed.*
			
D	S	Sie kontrollierte den $Ball_{amb}$, was sich im	**Tanz** mit dem Bräutigam deutlich zeigte.
		She controlled the $ball_{amb}$, which during the	***dance*** *with the bridegroom clearly showed.*
			
S	S	Sie kontrollierte den $Ball_{amb}$, was sich im	**Tanz** mit dem Bräutigam deutlich zeigte.
		She controlled the $ball_{amb}$, which during the	***dance*** *with the bridegroom clearly showed.*
			
S	D	Sie kontrollierte den $Ball_{amb}$, was sich im	**Spiel** beim Aufschlag deutlich zeigte.
		She controlled the $ball_{amb}$, which during the	***game*** *at the serve clearly showed.*
			

Introductory sentence was identical for all four conditions. The first two columns indicate the conveyed meaning of gesture and the subsequent target word: Dominant (D) or Subordinate (S). Target word is in **bold**. Literal translation is in *italics*. Cross-splicing was performed at the end of the main clause (i.e., in this case, after the word ''Ball'').

via a cross-splicing procedure (i.e., sentence parts from different recordings were combined using audio editing software). The aim of this procedure was to keep the speech in the dominant and the subordinate version of an item physically identical for as long as possible. For example, the sentence depicted in Table 1 was realized in the following way. The dominant sentence was used as it was recorded. The subordinate sentence was created by substituting the final part of the dominant sentence with the final part of a recording of the subordinate sentence (see Table 1 for more details). Thus, in this case, only the subordinate sentence was realized via cross-splicing. However, across the complete stimulus set, dominant and subordinate sentences were equally often cross-spliced.

The speech material was combined with the gesture videos resulting in a 2 × 2 design with gesture (Dominant vs. Subordinate) and target word (Dominant vs. Subordinate) as within-subject factors. The nylon stocking masked the mouth movements of the actress so

naïve participants could not identify the speech–lip mismatch in the two incongruent conditions. The final set consisted of 48 item quartets, resulting in a total of 192 sentences (see Table 1).

Rating of gesture phases. The onset of the gesture preparation as well as the on- and offset of the gesture stroke were independently assessed by two persons (interrater reliability >.90). These values did not differ significantly across gesture conditions [all $F(3, 94) < 1$] (for more information about the temporal relationship between gesture and speech in the experimental set, see Table 2).

Procedure

The participants were seated in a dimly lit, sound-attenuated chamber facing a computer screen. They were instructed to watch and listen carefully. Their task was to judge after each trial whether gesture and speech had been compatible. Note that in order to perform this task, participants had to compare the meaning indicated by the homonym–gesture combination in the initial part of the sentence with the meaning expressed by the target word in the following subclause. The videos were centered on a black background and extended for 10° visual angle horizontally and 8° vertically. A trial started with a fixation cross on the screen, which was presented for 2000 msec, followed by the video presentation. Immediately after the offset of the video, a question mark prompted the participants to respond. Feedback was only given if participants failed to respond within 2000 msec after the response cue. Response reaction times (RTs) were measured starting with the presentation of the question mark.

An experimental session (excluding time for electrode application) lasted approximately 90 min. The experiment had four blocks each consisting of 48 items. One block lasted approximately 8 min. A different, completely unrelated experiment was sandwiched between Blocks 1 and 2 (Part 1) and Blocks 3 and 4 (Part 2) to reduce memory strategies. The presentation order of the videos was varied in a pseudorandomized fashion, separately for each of the two parts. The order of the parts was reversed for half of the participants. In addition, the key assignment for correct (left or right) was counterbalanced across participants, resulting in a total of four experimental lists. One of the four lists was randomly assigned to each participant. That is, each list was seen by six participants.

ERP Recording

The EEG was recorded from 56 Ag/AgCl electrodes (Electrocap International). It was amplified using a PORTI-32/MREFA amplifier (DC to 135 Hz) and digitized on-line at 500 Hz. Electrode impedance was kept below 5 kΩ, and the left mastoid served as a reference. Vertical and horizontal electrooculograms (EOG) were also measured.

Data Analysis

Single-subject ERPs were calculated for each of the four conditions. The epochs were time-locked to the onset of the target word and lasted from 200 msec prestimulus onset to 1000 msec poststimulus onset. A 200-msec prestimulus baseline was used. Four regions of interest (ROIs) were defined: anterior-left (AL): AF7, AF3, F7, F5, F3, FT7, FC5, FC3; anterior-right (AR): AF4, AF8, F4, F6, F8, FC4, FC6, FT8; posterior-left (PL): TP7, CP5, CP3, P7, P5, P3, PO7, PO3; posterior-right (PR): CP4, CP6, TP8, P4, P6, P8, PO4, PO8. An automatic artifact rejection using a 200-msec sliding window was performed on the EOG channels (±30 μV) and on the EEG channels (±40 μV). Overall, approximately 30% of the trials did not enter statistical analysis due to artifacts or incorrect responses. Based on visual inspection of the data, a time window from 300 to 500 msec was used to analyze the N400 effects. The N400 is of crucial importance in the current paradigm to examine the impact of gesture on the integration of the target words. The potential effects after 500 msec were beyond the scope of this article and were therefore not statistically analyzed. Recall that the dominant and subordinate version of an item were matched up to the target word, which had a mean length of 380 msec (see Table 2). After the target word,

Table 2. Stimulus Properties

Gesture	*Target Word*	*Gesture Stroke Onset*	*Gesture Stroke Offset*	*Homonym Onset*	*Target Word Onset*	*Target Word Offset*
D	D	2.07 (0.46)	2.91 (0.48)	2.84 (0.40)	3.78 (0.38)	4.16 (0.38)
D	S	2.07 (0.46)	2.91 (0.48)	2.84 (0.40)	3.80 (0.38)	4.17 (0.38)
S	S	2.17 (0.52)	3.01 (0.51)	2.84 (0.40)	3.80 (0.38)	4.17 (0.38)
S	D	2.17 (0.53)	3.01 (0.51)	2.84 (0.40)	3.78 (0.38)	4.16 (0.38)
Mean		2.12 (0.49)	2.96 (0.50)	2.84 (0.40)	3.79 (0.38)	4.17 (0.38)

Mean onset and offset values are in seconds relative to the onset of the introductory sentence (*SD* in parentheses).

the dominant and subordinate sentences continued differently. Thus, the current design does not allow for a clear interpretation of effects occurring after the target word offset because such an effect could reflect an impact of the preceding gesture, the specific sentence continuation, or the interaction of both. For the ERP data, a repeated-measure analysis of variance (ANOVA) using the within-subject factors gesture (Dominant, Subordinate), target word (Dominant, Subordinate), part (1, 2), region (anterior, posterior), and hemisphere (left, right) was calculated. Only effects that involve the critical factors gesture target relation or target word meaning are reported. Greenhouse–Geisser (1959) correction was applied where necessary. In such cases, the uncorrected degrees of freedom (*df*), the corrected p values, and the correction factor ε are reported. Before entering statistical analysis, the data were filtered off-line with a high-pass filter of 0.2 Hz. For presentation purposes only, an additional 10-Hz low-pass filter was used.

Results

Behavioral Data

Performance was accurate for all four conditions (Dominant–Dominant [DD]: 92.0%; Subordinate–Subordinate [SS] 89.8%; Subordinate–Dominant [SD] 83.6%; Dominant–Subordinate [DS] 83.9%) and increased during the experimental run (Part 1: 84.3%; Part 2: 91.4%). An ANOVA with the factors gesture (2), target word (2), and part (2) revealed a significant three-way interaction between gesture, target word, and part [$F(1, 23) = 10.9$; $p < .0001$] as well as a two-way interaction between gesture and target word [$F(1, 23) = 35.66$; $p < .0001$]. Based on the three-way interaction, separate ANOVAs within each of the four conditions DD, DS, SD, and SS were carried out to analyze the simple main effects of part. The step-down analysis indicated that the increase in performance during the experimental run was only significant for conditions DS [$F(1, 23) = 36.63$; $p < .0001$], SD [$F(1, 23) = 10.68$; $p < .0001$], and SS [$F(1, 23) = 4.56$; $p < .05$], but not for condition DD [$F(1, 23) = 1.15$; $p > .29$]. To investigate the two-way interaction of Gesture by Target Word, the four conditions were compared via a series of Bonferroni-corrected post hoc tests. These tests indicated that the responses for condition DD were significantly more accurate than the responses for condition SD [$F(1, 23) = 42.44$; $p_{Bon} < .001$] and DS [$F(1, 23) = 43.21$; $p_{Bon} < .001$]. Similarly, the accuracy for condition SS was significantly greater than the accuracy for conditions SD [$F(1, 23) = 21.36$; $p_{Bon} < .001$] and DS [$F(1, 23) = 21.24$; $p_{Bon} < .001$]. No significant differences were observed between conditions DD and SS [$F(1, 23) = 4.81$; $p_{Bon} > .23$] as well as between conditions SD and DS [$F(1, 23) < 1$].

The RT showed a similar pattern in the four conditions (DD: 601 msec; SS 621 msec; SD 655 msec; DS 660 msec). The corresponding ANOVA showed only a two-way interaction between gesture and target word [$F(1, 23) = 13.57$; $p < .01$], indicating that the reaction time was longer for the incompatible conditions DS and SD as compared to the compatible conditions DD and SS.

ERP Data

As can be seen in Figure 1, the ERPs show an increased negativity for the incongruent conditions SD and DS starting around 300 msec. On the basis of its latency and scalp distribution, the negativity was identified as an N400. After the N400, a sustained negativity for in the incompatible conditions is visible at anterior sites. In addition, there is a positivity for condition SS at posterior sites.

To test the N400 effect, the mean amplitude from 300 to 500 msec time-locked to the onset of the target word was computed for all conditions. An ANOVA with the factors gesture (2), target word (2), part (2), region (2), and hemisphere (2) yielded significant two-way interactions between gesture and target word [$F(1, 23) = 30.64$; $p < .0001$] and between gesture and region [$F(1, 23) = 10.78$; $p < .01$]. The Gesture by Region interaction indicated that the N400 at target words following subordinate gestures was, in general, slightly more negative at anterior sites [$F(1, 23) = 3.64$; $p < .07$] but not at posterior sites [$F(1, 23) = 1.03$; $p > .32$]. Licensed by the Gesture by Target word interaction, the simple main effects of gesture at both types of target words were analyzed. At dominant target words, the N400 was larger after a subordinate gesture [$F(1, 23) = 26.23$; $p < .0001$]. Conversely, the N400 at subordinate target words was larger if preceded by a dominant gesture [$F(1, 23) = 18.44$; $p < .001$]. Thus, the activation of both the dominant and the subordinate word meaning varied reliably as a function of the congruency of the gesture context. The observed N400 effects were stable across the experimental run, that is, no significant interactions with the factor part were observed.

Discussion

The question addressed in Experiment 1 was whether dynamic co-speech gestures can be used as disambiguation cues. The behavioral data show that identifying an incongruent gesture–target word relation was associated with more errors and a longer RT. The ERP data show that the N400 at a dominant target word was larger after a subordinate gesture. Similarly, the N400 at a subordinate target word was larger following a dominant gesture.

The lower accuracy and higher RTs for conditions SD and DS suggest that identifying an incongruent relationship between the initial gesture–homonym context and the subsequent target word was a bit more difficult for

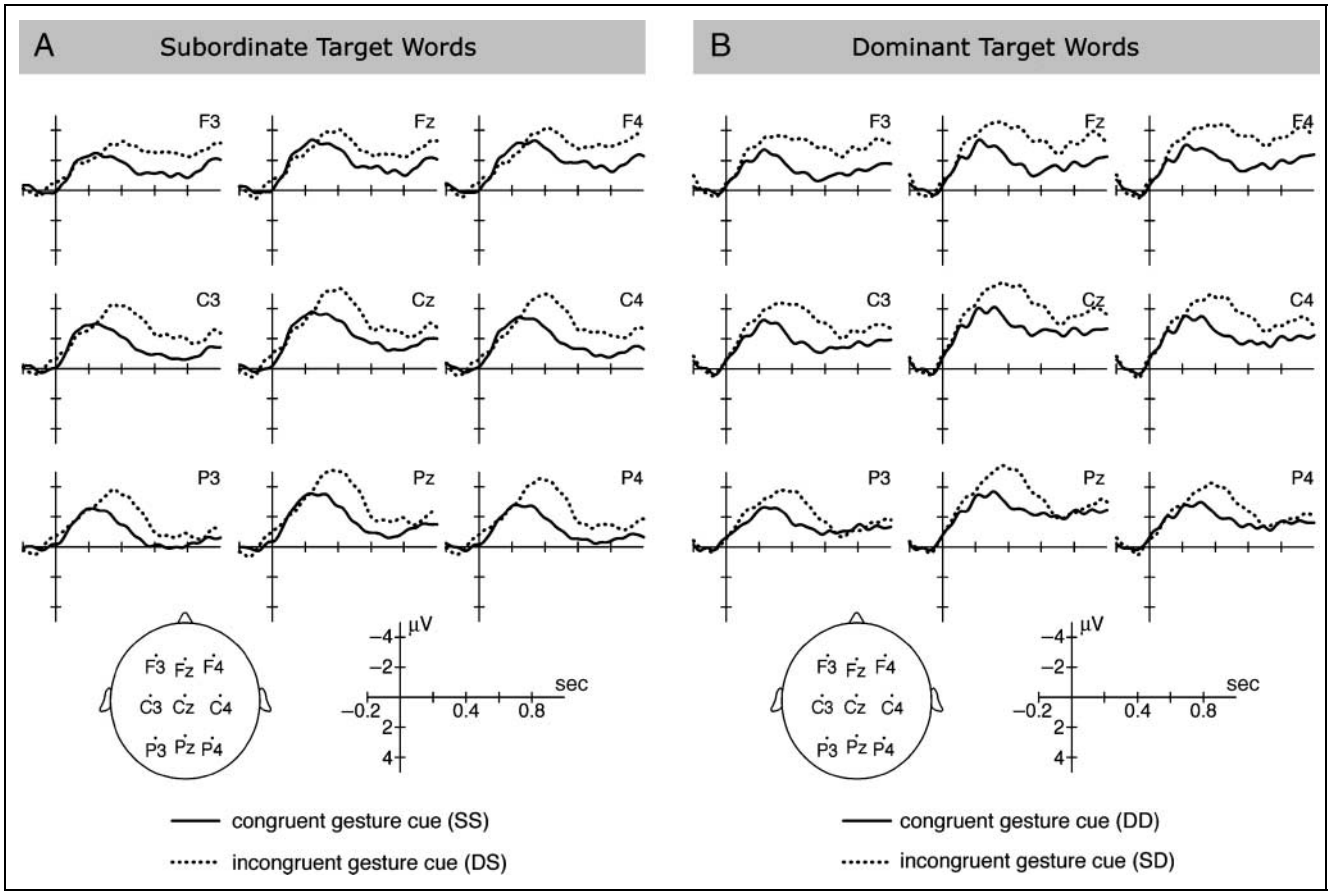

Figure 1. Pairwise presentation of the ERPs time-locked to the target word for all four conditions (n = 24 for each). Negativity is plotted up. In both the left and right panels, the solid line represents the instances in which the preceding nominal gesture cue and the target word were compatible. The dotted line represents the cases in which nominal gesture cue and target word were incompatible.

participants. It can be taken as initial evidence that participants used the gestural information for meaning selection, which caused some degree of interference at incongruent target words. However, because the congruency became already evident at the target word, but response was delayed until the offset of the video, the behavioral data should be interpreted with caution.

The ERP results indicate that the iconic gestures influenced the activation of the word meanings. The N400 at the subordinate target words was smaller after a subordinate gesture and larger after a dominant gesture. Thus, the subordinate meaning was more active in working memory after a subordinate gesture context and less active after a dominant gesture context (see Figure 1). Conversely, the N400 at dominant target words was smaller after dominant gesture and larger after a subordinate gesture. Thus, the dominant word meaning was more active in working memory after a dominant gesture and less active after a subordinate gesture. Taken together, the activation of both word meanings varied reliably as a function of the preceding gesture context. Because such a pattern of results is characteristic for strongly biasing context, we conclude that the iconic gestures constituted a strong contextual cue.

In summary, Experiment 1 demonstrated that listeners use gestural information to disambiguate speech. The pattern of results suggests that the gestures strongly biased the activation of the word meanings.

EXPERIMENT 2

One limitation of Experiment 1 might be the task that was employed. The participants had to compare the information from both gesture and speech, and explicitly judge their compatibility. Thus, the task forced participants to combine gesture and speech. Experiment 2 investigates whether iconic gestures are still used as disambiguating cues once the task is less explicit and no longer requires an integration between gesture and speech.

Methods

Participants

In Experiment 2, 25 native German speakers participated, none of whom had participated in Experiment 1. One subject had to be excluded due to excessive artifacts. The

remaining 24 participants (12 women) had a mean age of 25 (range 21–29 years) and were right-handed (laterality coefficient = 95; Oldfield, 1971). All participants had normal or corrected-to-normal vision, and none reported any known hearing deficit.

Stimuli

The same stimuli as in Experiment 1 were used.

Procedure

Presentation of stimuli was identical to Experiment 1, however, participants were instructed to perform a different task. The aim of the task was to ensure that participants attended to both the visual and auditory streams of the video. However, the task should not require participants to combine both streams of information and give no cue as to how the arm movements might be related to the contents of speech. The participants received the following instructions: ''In this experiment, you will be seeing a number of short videos with sound. During these videos the speaker moves her arms. After some videos, you will be asked whether you have seen a certain movement or heard a certain word in the previous video.''

A visual prompt cue was presented after the offset of each video. After 87.5% of all videos, the prompt cue indicated the upcoming trial (i.e., no response was required in these trials).

After 6.25% of all videos, the prompt cue asked participants to prepare for the movement task. A short silent video clip was presented as a probe. The probes were taken from the experimental stimuli and only contained that portion during which the arm movement was executed. After the offset of the probe video, a question mark prompted the participants to respond. Feedback was given if participants answered incorrectly or if they failed to respond within 2000 msec after the response cue. RTs were measured starting with the presentation of the response cue.

After the remaining 6.25% of the videos, the prompt cue informed the participants that the word task had to be performed. Participants had to indicate whether a visually presented probe word had been included in the previous sentence. The probe words were selected from sentence-initial, -middle, and -final positions of the complex sentence. Response and feedback were identical to the movement task trials.

ERP Recording and Data Analysis

The parameters for the recording and the analysis of the data were the same as in Experiment 1. Because the movement task was performed after only 6.25% of all trials, we obtained very few gesture-related behavioral data (i.e., only four responses per condition). Therefore, we do no report a statistical analysis of the behavioral data. Behavioral responses were also not a rejection criterion for the ERP trials. Based on the artifact rejection, approximately 11% of the trials were excluded from statistical analysis. As in Experiment 1, a time window ranging from 300 to 500 msec was selected based upon visual inspection of the data for the statistical analysis of the N400 effects.

Results

ERP Data

An enhanced N400 for the incompatible conditions SD and DS is visible starting around 300 msec (see Figure 2). In addition, the ERPs for subordinate target words appear more negative as compared to dominant target words.

The ANOVA for the time window from 300 to 500 msec revealed a significant two-way interaction between gesture and target word [$F(1, 23) = 15.46$; $p < .001$] and a main effect of target word [$F(1, 23) = 7.84$; $p < .05$]. The main effect of target word indicated that the N400 was slightly more negative at subordinate target words. On the basis of the two-way interaction, the simple main effects of gesture were tested separately. At dominant target words, the N400 was larger after subordinate gesture [$F(1, 23) = 4.72$; $p < .05$]. Similarly, at the subordinate target words, the N400 was larger after a dominant gesture [$F(1, 23) = 10.32$; $p < .01$]. Thus, both word meanings varied reliably as a function of the preceding gesture context. The N400 effects did not interact with the factor part, and thus, were stable across the experimental run.

Discussion

The aim of Experiment 2 was to clarify the extent to which the results obtained in Experiment 1 were task-dependent. As in Experiment 1, broadly distributed N400 effects with similar latencies were observed for the incongruent conditions DS and SD as compared to the congruent conditions SS and DD. This is in analogy to the discussion of Experiment 1 interpreted to reflect a strong impact of the gesture context on disambiguation. Thus, iconic gestures are used as disambiguation cues, even if disambiguation is not explicitly required by the task.

In contrast to the results of Experiment 1, there appears to be a general negative-going trend for all conditions. To investigate this negative shift, we extracted longer ERPs, which included the prompt cue that was presented about 1400 msec after target word onset. These prolonged ERPs show a slowly rising and frontally distributed negativity peaking at 1600 msec. After the

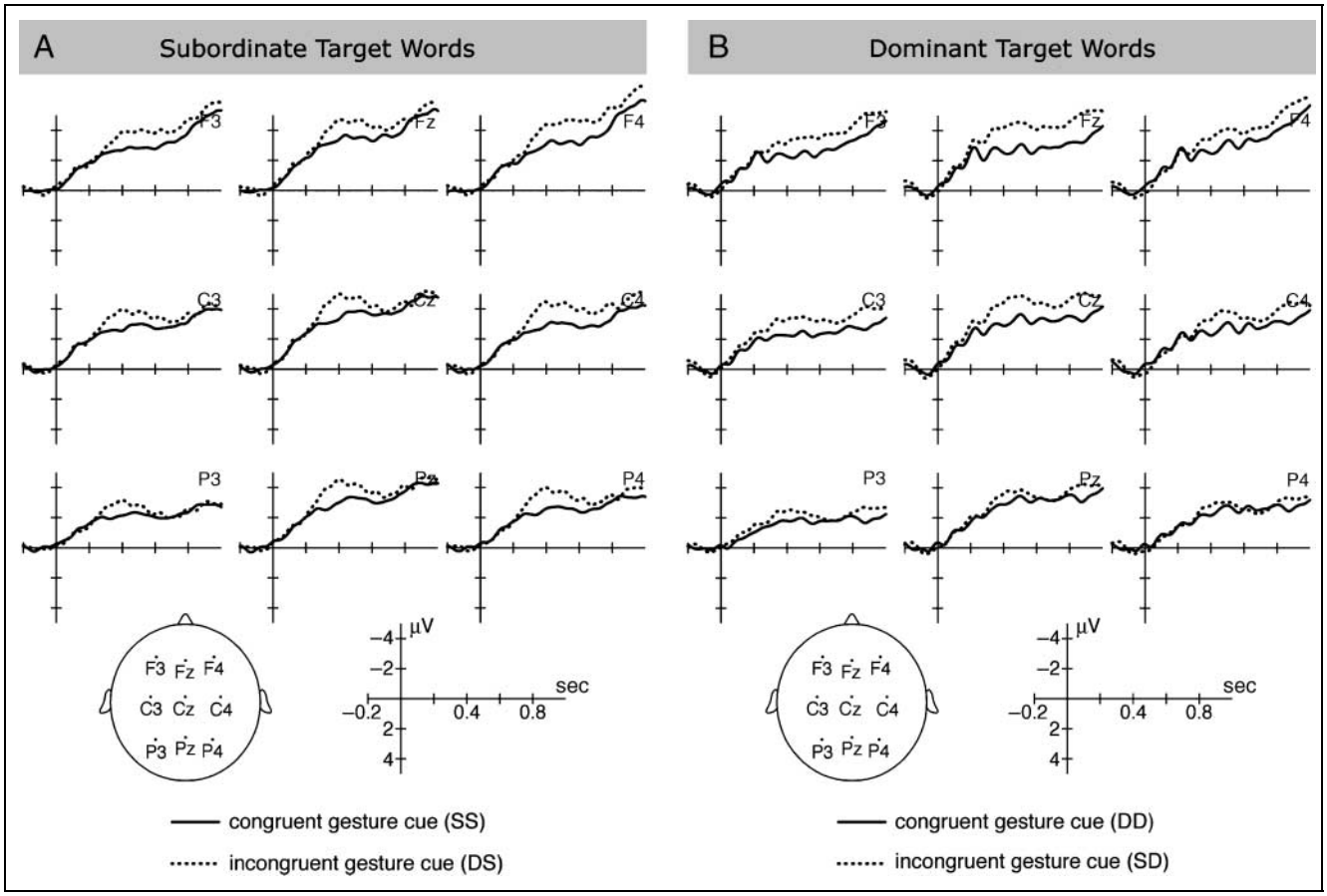

Figure 2. Pairwise presentation of the ERPs time-locked to the target word for all four conditions (n = 24 for each). The solid line represents the instances in which the preceding nominal gesture cue and the target word were compatible, the dotted line represents the incompatible instances.

peak, the ERPs for all conditions clearly return to baseline level. Based on the scalp distribution and the task manipulation, we interpret this negative shift as a contingent negative variation (CNV, see, for instance, Rugg & Coles, 1995). In Experiment 2, the participants did not know whether a task was coming up or not until the offset of the video. It is therefore not surprising that the participants built up expectations about the upcoming prompt cue during the final part of the gesture clip. This general expectation is suggested to be reflected in a CNV.

In sum, the results from Experiments 1 and 2 are in line with previous ERP studies (Wu & Coulson, 2005; Kelly et al., 2004) in showing that an initial gesture context can modulate the processing of a subsequent target word. The present study extends the previous findings in showing that contextual effects of gesture are not restricted to the processing of isolated target words but can be generalized to auditory sentence processing.

EXPERIMENT 3

Experiment 1 and Experiment 2 have shown that speakers use gestural information to disambiguate speech. The results from Experiment 2 suggest that this disambiguating effect is somewhat task-independent. This can be taken as some initial evidence that a listener performs an obligatory integration of gesture and speech as discussed by McNeill and coworkers (1994). However, as has been mentioned in the Introduction, another important test for the potential automatic nature of gesture–speech integration is the addition of meaningless hand movements. As we all know, apart from producing gestures, speakers also use their hands to make conversationally irrelevant, meaningless movements while engaged in a conversation (e.g., a speaker might scratch his or her chin, rub his or her temple, squeeze his or her nose, and so on). These grooming movements (also called adaptors or manipulators) are typically very repetitive and the speaker is hardly aware of them (Goldin-Meadow, 2003). People probably differ not only in their individual set of grooming movements but also in the frequency with which they exhibit these behaviors (Ekman, 1999). The impact of grooming on comprehension is not well understood, although there is some evidence that excessive grooming movements can cause a speaker to appear less trustworthy (DePaulo et al., 2003). An important difference between gesture

and grooming is that grooming is not systematically tied to a segment of speech. Thus, a grooming movement in the current disambiguation paradigm gives the listener no cue for meaning selection.

One advantage of grooming, therefore, is that it constitutes a neutral context, which allows to investigate whether the effects of gesture are inhibitory or facilitatory in nature. In Experiments 1 and 2, it was observed that the N400 at subordinate target words was larger after a dominant gesture and smaller after a subordinate gesture. Without an unrelated condition, it is impossible to tell whether this effect occurred because the dominant gesture inhibited the subordinate meaning or because the subordinate gesture actually facilitated the subordinate meaning.

Another advantage of adding meaningless hand movements to the paradigm is that it allows us to test whether gesture–speech integration is a primarily automatic process. Adding the grooming movements makes the manual modality less informative because, under such circumstances, only a portion of all observed hand movements provide the listener with a helpful cue for disambiguation. If the integration of gesture and speech is a primarily automatic process, the addition of meaningless hand movements should not weaken the disambiguating effects of gesture, that is, the N400 of both the dominant and the subordinate target words should vary as a function of the congruency of the preceding gesture context, as it was the case in Experiments 1 and 2. If, however, the integration of gesture and speech is also substantially influenced by controlled factors, a different pattern of results should emerge. A listener may start to consider *all* manual cues (including the gestures) as less informative once meaningless hand movements are added. Such a devaluation of gesture could result in a pattern typical for weakly biasing contexts (i.e., only the N400 at subordinate target words would vary as a function of the preceding gesture). Finally, it is possible that listeners completely disregard the semantic content of gesture once grooming is added. In this case, the N400 at the target words should vary only as a function of word meaning frequency. Based on the results of Experiments 1 and 2 as well as the data from Özyürek et al. (2007) and Kelly et al. (2004), we hypothesize that gesture–speech integration is an obligatory process, that is, the addition of meaningless grooming movements should not weaken the impact of gesture on homonym disambiguation.

Methods

Participants

Twenty-nine subjects participated in Experiment 3, none of whom had participated in Experiment 1 or 2. Five participants did not enter statistical analysis because of excessive artifacts. The remaining 24 participants (12 women) had a mean age of 24 years (range 19–28 years) and were right-handed (laterality coefficient = 93; Oldfield, 1971). All participants had normal or corrected-to-normal vision, and none reported any known hearing deficit.

Stimuli

In addition to the stimuli used in Experiments 1 and 2, a third gesture condition was added. Again, our professional actress was videotaped while uttering the sentence stimuli. Instead of the disambiguating gestures, she performed meaningless grooming hand movements (scratching, rubbing, etc.). The sentence material described in Experiment 1 was combined with the newly recorded material, resulting in a 3 × 2 design with gesture (**D**ominant, **S**ubordinate, **G**rooming) and target word (**D**ominant, **S**ubordinate) as within-subject factors (see Table 3). The three types of hand movements did not differ significantly in the onset and offset of the movement stroke [both $F(2, 141) < 1$] (Table 4).

Procedure

The less explicit task from Experiment 2 and the same instructions were employed. The experiment consisted of six blocks consisting of 48 items, in which each block lasted approximately 7 min. In contrast to Experiments 1 and 2, there was no unrelated second experiment embedded at half time, thus the total time of the experimental session (excluding time for electrode application) was reduced to 50 min. For the statistical analysis, the data of each participants were divided into three parts (Part 1: Blocks 1 and 2; Part 2: Blocks 3 and 4; Part 3: Blocks 5 and 6). Two pseudorandomized lists were created. The key assignment for correct (left or

Table 3. Examples of Additional Stimuli Used in Experiment 3

Table 4. Properties of Additional Stimuli

Gesture	*Speech*	*Gesture Stroke Onset*	*Gesture Stroke Offset*	*Homonym Onset*	*Target Word Onset*	*Target Word Offset*
G	D	2.16 (0.49)	2.96 (0.50)	2.84 (0.40)	3.78 (0.38)	4.16 (0.38)
G	S	2.16 (0.49)	2.96 (0.50)	2.84 (0.40)	3.80 (0.38)	4.17 (0.38)

Mean onset and offset values are in seconds relative to the onset of the introductory sentence (*SD* in parentheses).

right) was also balanced across participants, resulting in a total of four experimental lists.

ERP Recording and Data Analysis

The data were amplified using a BrainAmp MR plus amplifier (DC to 250 Hz). For the ERP data, a repeated-measure ANOVA using the within-subject factors gesture (D, S, G), target word (D, S), region (anterior, posterior), hemisphere (left, right), and part (1, 2, 3) was calculated. Based on the artifact rejection, approximately 14% of trials were excluded from statistical analysis. The time window for the statistical analysis of the N400 effect was set from 300 to 500 msec based on visual inspection of the data. All other recording and analysis details were as described in Experiment 1.

Results

ERP Data

As can be seen in Figure 3, the processing of subordinate target words is associated with a larger N400 if preceded by a incongruent dominant gesture or an unrelated grooming movement. The ERPs for these conditions (DS and GS) remain more negative than condition SS even after the N400 time window. At the dominant target words, there may be anteriorly an increased negativity

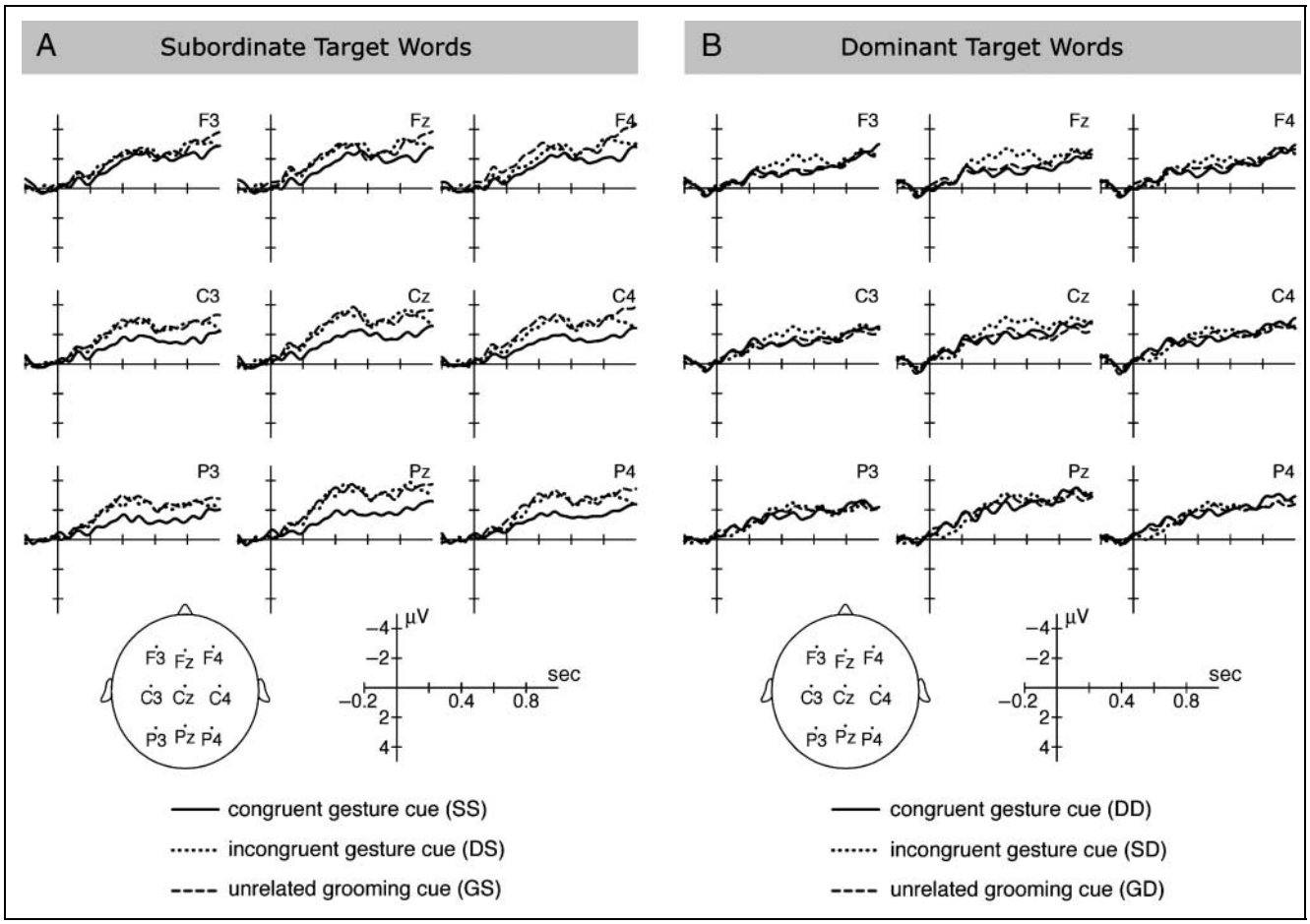

Figure 3. Presentation of the ERPs time-locked to the target word in two sets of three conditions (n = 24 for each). The solid line represents the instances in which the preceding nominal gesture cue and the target word were compatible, the dotted line represents the incompatible instances. The dashed line represents those cases in which a target word was preceded by an unrelated grooming movement.

for dominant targets following a subordinate gesture. Finally, there is a general negative-going trend for all conditions, especially at anterior sites.

An ANOVA for the time window from 300 to 500 msec yielded a significant three-way interaction between target word, part, and region [$F(2, 46) = 4.1$; $p < .05$; $\varepsilon = .95$], a marginally significant two-way interaction between gesture and target word [$F(2, 46) = 3.28$; $p < .06$; $\varepsilon = .81$], as well as a main effect of target word [$F(1, 23) = 7.16$; $p < .05$].

A number of step-down analyses were performed to clarify the origin of the three-way interaction. In a first step, separate ANOVAs with the factors target word and region were calculated within each level of part. There were no significant effects or interactions in Part 1 [all $F(1, 23) < 2.03$; all $p > .17$] or in Part 3 [all $F(1, 23) < 1$]. However, it was found that in Part 2 of the experiment, the N400 was significantly larger at subordinate target words [$F(1, 23) = 7.02$; $p < .05$]. This was especially the case at anterior sites, as indicated by a significant Target word by Region interaction [$F(1, 23) = 4.91$; $p < .05$]. Thus, there was a significant deviation in the way the subordinate target words were processed during Part 2 of the experiment. Note, however, that the interaction did not involve the factor gesture, which is of crucial interest in the current design.

The main effect of target word was found to indicate a larger N400 at subordinate words. Based on the two-way interaction of Gesture by Target word, the simple main effects of gesture at both types of target words were analyzed. The simple main effect of gesture was significant at subordinate target words [$F(2, 46) = 4.56$; $p < .01$; $\varepsilon = .88$], but not at dominant targets [$F(2, 46) < 1$]. Finally, the three levels of gesture at subordinate targets were contrasted via three Bonferroni-corrected post hoc tests. It was found that the N400 at subordinate target words was larger after grooming than after a subordinate gesture [$F(1, 23) = 6.99$; $p_{Bon} < .05$] and larger after a dominant gesture than after a subordinate gesture [$F(1, 23) = 6.93$; $p_{Bon} < .05$]. The difference between grooming and dominant gestures at subordinate targets was not significant [$F(1, 23) < 1$].

On the basis of the two-way interaction of Gesture by Target word, the simple main effect of speech was analyzed for grooming. It was found that the N400 at subordinate targets following grooming was more negative than the N400 at dominant targets following grooming [$F(1, 23) = 9.96$, $p < .05$]. Thus, in the absence of a contextual cue, word meaning frequency determined the activation pattern.

Because it was found that the N400 at dominant target words did not vary as a function of context congruency, one immediate follow-up question concerned the extent to which the dominant meaning was activated at all in Experiment 3. To address this issue, the N400 amplitude of the conditions with a dominant target word (DD, GD, SD) was compared to a condition where the gesture context had clearly caused a higher activation of the contextually appropriate word meaning (condition SS). The corresponding ANOVA was not significant [$F(3, 69) < 1$], suggesting that the dominant word meaning was activated in Experiment 3.

As in Experiment 2, long epochs ranging from −200 to 2000 msec relative to target word onset were extracted to clarify the nature of the negative-going trend. Again, an anteriorly distributed and slowly rising negativity with its peak at 1600 msec was revealed. After the peak, the negativity quickly returned to the baseline level. Thus, the negative-going trend is suggested to reflect a CNV as in Experiment 2.

GENERAL DISCUSSION

The present series of experiments investigated whether listeners use the information from iconic gestures to disambiguate unbalanced homonyms. In Experiment 1, participants were explicitly asked to judge the compatibility between an initial homonym–gesture combination and a subsequent target word. ERPs time-locked to the target word revealed that the N400 was smaller after a congruent gesture context and larger after an incongruent gesture context, suggesting that listeners can use gestural information to disambiguate speech. Experiment 2 replicated the results using a less explicit task, indicating that the disambiguating effect of gesture is somewhat task-independent. Unrelated grooming movements were added to the paradigm in Experiment 3. This manipulation changed the pattern of results. Only the N400 at the subordinate target words varied as a function of the preceding gesture context, whereas the N400 at dominant target words did not.

In Experiment 3, the activation of the subordinate meaning varied as a function of context congruency just as it was observed in Experiments 1 and 2. Grooming as an unrelated condition allows for a clearer interpretation of the N400 effect. Both grooming as well as a dominant gesture context make the integration of a subsequent subordinate target word more difficult as reflected by the increased N400. Only a subordinate gesture context leads to an attenuated N400 at subordinate target words. This result suggests that the underlying mechanism is facilitatory and not inhibitory because the N400 is smaller after a congruent subordinate gesture than after a neutral grooming context. Thus, a gesture that supports the lesser frequent meaning of a homonym actually facilitates the processing of a related target word in on-line sentence comprehension. This is an important extension of the existing ERP literature on iconic gesture comprehension (Özyürek et al., 2007; Wu & Coulson, 2005; Kelly et al., 2004) because previously conducted ERP studies on iconic gesture comprehension have only demonstrated how gesture can impair speech processing.[2] The present study shows that disambiguation is one mechanism

through which gesture information can also *facilitate* speech comprehension. A listener can save resources by using the gestural information to activate the lesser frequent meaning of a homonym. Perhaps it is also possible to divert the saved resources to other sources of information (e.g., prosody, body posture), which would have been missed without the facilitatory effect of gesture. However, this is subject to further research.

Because it has been frequently observed that the N400 is sensitive to the degree to which a target word fits into a given context (Kutas & Federmeier, 2000), one could have expected that the N400 at subordinate target words would be larger after an incongruent dominant gesture context than after a neutral grooming context. However, the size of the N400 effects that grooming and dominant gestures elicited at subordinate target words did not differ. One possible explanation is that this is due to a floor effect. Given that the target word is processed (on average) some 900 msec after the homonym/hand-movement combination, it may very well be that in the case of grooming, the subordinate meaning is no longer active at all at the position of the target word. Simpson and Burgess (1985) reported data suggesting that in a neutral context, the subordinate meaning is maintained only for 500 msec following homonym presentation. In light of such data, the similar N400 effects at subordinate target words may reflect that after the processing of both types of hand movements (i.e., grooming and dominant gesture), the subordinate meaning is completely deactivated by the time the subordinate target word is encountered.

There were no significant N400 differences at the dominant target word in Experiment 3. The N400 for the three context types at dominant target words (conditions DD, GD, and SD) did not differ significantly from the N400 for condition SS, suggesting that the dominant word meaning was activated after all three types of hand movements. Taken together, it was found in Experiment 3 that the activation of the subordinate meaning varied reliably as a function of context congruency, whereas the dominant meaning was equally active after a congruent dominant gesture as well as after an incongruent subordinate gesture. As has been mentioned in the Introduction, such pattern of results is typical for weakly biasing context cues. It seems that once a listener is confronted with a mixture of meaningless grooming movements and meaningful gestures, the impact of gesture on speech disambiguation is weakened. Thus, the integration of gesture and speech in comprehension is not a purely automatic process but is modulated by situational factors, in this case, the proportion of meaningful and meaningless hand movements. In the following, we discuss some possible mechanisms through which the addition of grooming may have weakened the impact of gesture.

It is, in principle, possible that the cause for the weaker impact of gesture is simply the increased number of homonym repetitions in Experiment 3. For example, participants may initially have taken gesture into account, but as they became more familiar with the stimulus set during the experiment, they may have ceased to pay attention to gestures, as they realized that the gestures actually provided no helpful cue. In terms of statistical factors, such a scenario would imply an interaction between the factors gesture and part. However, such an interaction was not found in the statistical analysis. The only significant interaction involving the factor part was a three-way interaction between target word, region, and part (see Results section), which cannot explain the different pattern of results in Experiment 3.

Another possible explanation is that the processing of grooming movements interfered with the processing of the gestures. It is conceivable that the participants misinterpreted some grooming movements as gestures on the one hand and mistook some gestures as grooming on the other hand. The outcome of such misinterpretations would be a weaker impact of gesture. However, we think such an ''active interference'' explanation is not very likely for at least two reasons. First, grooming caused the expected neutral context pattern, that is, the N400 was smaller at dominant targets and larger at subordinate targets after a grooming movement. This result suggests that grooming was an acceptable neutral context for the participants with respect to meaning selection and not a strange distracting movement that interfered with speech processing. Second, in a currently conducted study (Holle, Gunter, Rüschemeyer, Hennenlotter, & Iacoboni, under revision), we asked participants after the experiment whether it is difficult to distinguish grooming from gesture. Participants very consistently report that it is quite easy to make this distinction because a gesture is usually a more dynamic movement than grooming.

An alternative explanation for the weaker impact of gesture in Experiment 3 is that the addition of grooming decreased the degree to which listeners took gestural information into account. In this experiment, there was only a 66% chance that an observed hand movement conveyed meaning. This reduced probability may have caused listeners to put less weight on gestural information and more weight on other sources of information (in this case, word meaning frequency) during the meaning selection process. If this explanation is valid, an interesting question for future research will be how much meaningless hand movements are tolerable for a listener before the impact of gesture is weakened. The present study used a 2:1 ratio of meaningful versus meaningless hand movements. Although the ratio in natural face-to-face conversation is unknown, it may very well be the case that listeners are used to seeing a higher proportion of meaningful movements (e.g., 3:1).

Whatever the underlying mechanism, it is clear that the addition of grooming weakened the impact of

gesture. This result is incompatible with the automaticity notion of gesture–speech integration from McNeill (''the point we wish to emphasize is the involuntary, automatic character of forming an idea unit out of information from the two channels''; Mc Neill et al., 1994, p. 236). The present study suggests that gesture–speech integration does not operate in such a modular fashion. Instead, external factors such as the proportion of meaningful to meaningless hand movements can also influence the degree to which listeners take gesture into account. This finding may have implications for research on gesture–speech production. For example, it might be the case that speakers who produce few to no grooming movements are more effective communicators than speakers with a high individual grooming frequency. Our finding that gesture–speech integration is not an entirely automatic process is also in line with recent data by Kelly et al. (in press), who suggested that the intentional relationship between gesture and speech also influences the degree to which both channels of information interact in comprehension.

Another factor that seems to moderate the impact of gesture is the amount of semantic overlap between gesture and speech. Özyürek et al. (2007) found that in a clearcut mismatch with no semantic overlap between gesture and speech (e.g., gesturing *rolling down* while saying *knock*), the processing of speech is negatively affected as indexed by an enlarged N400. Kelly et al. (2004) realized different degrees of semantic overlap in their experiment. Gesture and speech were either highly overlapping (match), partially overlapping (complementary), or not overlapping (mismatch). In this study, the processing of target words that mismatched the preceding gesture was associated with an enlarged N400 as compared to the match or complementary condition. In neither of the two studies did the task require taking gesture into account, suggesting that the interference caused by semantically nonoverlapping gesture–speech combinations was inevitable in these experiments. The present study contained stimuli with a moderate semantic overlap. Gesture and speech were semantically overlapping in that they both referred to the same homonym. However, both domains were also nonoverlapping, in that gesture always conveyed additional information about the appropriate word meaning of the homonym. In Experiment 3, it was found that such a moderate amount of semantic overlap can facilitate the processing of the lesser frequent word meaning. We suggest that the amount of semantic overlap determines whether an iconic gesture has an interfering or a facilitating effect on comprehension. If there is almost no overlap between gesture and speech, as in the case of clearcut mismatches, gesture has an interfering effect on comprehension. If there is a moderate amount of semantic overlap, as in the case of disambiguation, gesture can facilitate speech comprehension.[3] The idea that a listener benefits from a moderate semantic overlap between gesture and speech is also in line with behavioral data from Goldin-Meadow and Momeni-Sandhofer (1999) and Alibali, Flevares, and Goldin-Meadow (1997).

The present study demonstrated that the integration of gesture and speech in comprehension can be modulated by situational factors such as the amount of meaningful hand movements in an experiment. It would be interesting to see whether the addition of meaningless hand movements also weakens the impact of gesture in a mismatch paradigm. Such an experiment would allow some conclusions about which of the two factors—semantic overlap or proportion of meaningful hand movements—plays a more prominent role in determining the degree to which gesture is taken into account.

Conclusion

In sum, there are two main conclusions from the present study. First, listeners use gestural information to disambiguate speech. Particularly, the processing of a lesser frequent meaning of a homonym can be facilitated by iconic gestures. Second, the integration of gesture and speech is not an obligatory process, but is modulated by situational factors. Once the listener is confronted with a mixture of meaningful and meaningless hand movements, the impact of gesture is weakened.

Acknowledgments

We thank Susanne Wagner for supplying the original material and for useful discussions, Bettina Johst for her programming efforts, Shirley-Ann Rüschemeyer for editing the article from a native-speaker perspective, Kristiane Werrmann and Christiane Hoffmann for data acquisition, Sven Gutekunst for technical assistance, and Korinna Eckstein and Anna Hasting for their help during the selection of the gestures. Additionally, we are grateful to three anonymous reviewers for their helpful comments on an earlier draft of this article.

Reprint requests should be sent to Henning Holle, Max Planck Institute for Human Cognitive and Brain Sciences, PO Box 500 355, 04303 Leipzig, Germany, or via e-mail: holle@cbs.mpg.de.

Notes

1. In this overview on the literature on homonym processing, we refrain from absolute statements about the activation of the two word meanings, such as *the dominant meaning is active, the subordinate meaning is not active*. Instead, we focus on whether the activation of a word meaning varies as a function of the preceding context (e.g., *The dominant meaning is more active after a dominant context than after a subordinate context*). The rationale of this focus is to facilitate the comparability between the literature and the present study.
2. It should be noted that Wu and Coulson (2005) discussed their findings as a facilitatory effect of iconic gestures. Their paradigm, however, did not focus on co-speech gestures.
3. The data from Kelly et al. may appear incompatible with this suggestion because, in that study, the N400 in the complementary condition (with a moderate gesture–speech overlap) was not more attenuated as compared to the no-gesture condition. Note, however, the possibility that a facilitatory effect

of gesture was not observed in this study because of a potential ceiling effect. Because each of the four target words was repeated 48 times during that experiment, the N400 for the target words may already have been attenuated to a large degree during the no-gesture condition, leaving ''no room'' for an additional facilitatory effect of gesture in the complementary condition.

REFERENCES

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

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

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. (2001). An experimental investigation of the role of different types of iconic gesture in communication. *Gesture, 1,* 129–149.

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,* 179–192.

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

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129,* 74–118.

Ekman, P. (1999). Emotional and conversational nonverbal signals. In L. Messing & R. Campbell (Eds.), *Gesture, speech and sign* (pp. 45–55). London: Oxford University Press.

Feyereisen, P., Van de Wiele, M., & Dubois, F. (1988). The meaning of gestures: What can be understood without speech? *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 8,* 3–25.

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

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

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

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

Gunter, T. C., Wagner, S., & Friederici, A. D. (2003). Working memory and lexical ambiguity resolution as revealed by ERPs: A difficult case for activation theories. *Journal of Cognitive Neuroscience, 15,* 643–657.

Hadar, U., & Pinchas-Zamir, L. (2004). The semantic specificity of gesture—Implications for gesture classification and function. *Journal of Language and Social Psychology, 23,* 204–214.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11,* 194–205.

Hinojosa, J. A., Martin-Loeches, M., & Rubia, F. J. (2001). Event-related potentials and semantics: An overview and an integrative proposal. *Brain and Language, 78,* 128–139.

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

Holle, H., Gunter, T. C., Rüschemeyer, S. A., Hennenlotter, A., & Iacoboni, M. (under revision). Neural correlates of the processing of co-speech gestures.

Holler, J., & Beattie, G. (2003). Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener? *Gesture, 3,* 127–154.

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

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

Koyama, S., Nageishi, Y., & Shimokochi, M. (1992). Effects of semantic context and event-related potentials: N400 correlates with inhibition effect. *Brain and Language, 43,* 668–681.

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

Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991, November). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology, 61,* 743–754.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4,* 463–470.

Levelt, W. J., Richardson, G., & la Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language, 24,* 133–164.

Martin, C., Vu, H., Kellas, G., & Metcalf, K. (1999). Strength of discourse context as a determinant of the subordinate bias effect. *Quarterly Journal of Experimental Psychology, Series A, 52,* 813–839.

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

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

McNeill, D., Cassell, J., & McCullough, K.-E. (1994). Communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction, 27,* 223–237.

Nobe, S. (2000). Where do most spontaneous representational gestures actually occur with respect to speech? In D. McNeill (Ed.), *Language and gesture* (pp. 186–198). Cambridge: Cambridge University Press.

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

Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition, 9,* 225–236.

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

Paul, S. T., Kellas, G., Martin, M., & Clark, M. B. (1992). Influence of contextual features on the activation of ambiguous word meanings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 703–717.

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

Rugg, M. D., & Coles, M. G. H. (1995). Event-related brain potentials: An introduction. In M. D. Rugg & M. G. H. Coles

(Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition.* Oxford: Oxford University Press.
Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information-processing: 1. Detection, search, and attention. *Psychological Review, 84,* 1–66.
Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information-processing: 2. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.
Simpson, G. B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior, 20,* 120–136.
Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 28–39.
Simpson, G. B., & Krueger, M. A. (1991). Selective access of homograph meanings in sentence context. *Journal of Memory and Language, 30,* 627–643.
Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language, 27,* 324–340.
Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research, 49,* 161–167.
Twilley, L. C., & Dixon, P. (2000). Meaning resolution processes for words: A parallel independent model. *Psychonomic Bulletin & Review, 7,* 49–82.
Vu, H., Kellas, G., & Paul, S. T. (1998). Sources of sentence constraint on lexical ambiguity resolution. *Memory & Cognition, 26,* 979–1001.
Wagner, S. (2003). *Verbales Arbeitsgedächtnis und die Verarbeitung lexikalisch ambiger Wörter in Wort- und Satzkontexten* (Ph.D. thesis). Leipzig: Max-Planck-Institute for Cognitive Neuroscience.
Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology, 42,* 654–667.