Memory Effects of Speech and Gesture Binding: Cortical and Hippocampal Activation in Relation to Subsequent Memory Performance

Benjamin Straube[1], Antonia Green[1], Susanne Weis[1], Anjan Chatterjee[2], and Tilo Kircher[1]

Abstract

■ In human face-to-face communication, the content of speech is often illustrated by coverbal gestures. Behavioral evidence suggests that gestures provide advantages in the comprehension and memory of speech. Yet, how the human brain integrates abstract auditory and visual information into a common representation is not known. Our study investigates the neural basis of memory for bimodal speech and gesture representations. In this fMRI study, 12 participants were presented with video clips showing an actor performing meaningful metaphoric gestures (MG), unrelated, free gestures (FG), and no arm and hand movements (NG) accompanying sentences with an abstract content. After the fMRI session, the participants performed a recognition task. Behaviorally, the participants showed the highest hit rate for sentences accompanied by meaningful metaphoric gestures. Despite comparable old/new discrimination performances (d') for the three conditions, we obtained distinct memory-related left-hemispheric activations in the inferior frontal gyrus (IFG), the premotor cortex (BA 6), and the middle temporal gyrus (MTG), as well as significant correlations between hippocampal activation and memory performance in the metaphoric gesture condition. In contrast, unrelated speech and gesture information (FG) was processed in areas of the left occipito-temporal and cerebellar region and the right IFG just like the no-gesture condition (NG). We propose that the specific left-lateralized activation pattern for the metaphoric speech–gesture sentences reflects semantic integration of speech and gestures. These results provide novel evidence about the neural integration of abstract speech and gestures as it contributes to subsequent memory performance. ■

INTRODUCTION

The ability to classify entities or events as old or new is a fundamental function of everyday life. In communication, remembering information in an ongoing dialog is critical for effective discourse. Human speech is typically accompanied by gestures. These gestures confer advantages in comprehension and memory of spoken sentences (e.g., Valenzeno, Alibali, & Klatzkya, 2003; Kelly, Barr, Church, & Lynch, 1999). However, little is known about the neural processes underlying the integration/binding of speech and gesture information, and particularly their relation to subsequent memory performance.

Memories for complex events include a wealth of information. Encoding such multimodal experiences not only requires memory for particular features, but also cognitive processes that bind these features together. Theories about the neural bases of memory binding have focused mainly on interactions between the hippocampus and cortical regions involved in the on-line processing of a stimulus (e.g., Norman & O'Reilly, 2003; Shastri, 2002; Rolls, 2000; Alvarez & Squire, 1994; Eichenbaum, Otto, & Cohen, 1992; Marr, 1971). On these accounts, an event is represented in terms of the pattern of cortical activity engendered when it was experienced initially. Components of an event that were processed and represented in different cortical regions are thought be bound into a common memory representation by the hippocampus. Evidence supporting a hippocampal role in memory binding comes from nonhuman and human lesion studies, as well as from functional neuroimaging experiments (see Eichenbaum, 2004 for a review).

Several models of memory function (e.g., Norman & O'Reilly, 2003; Shastri, 2002; Rolls, 2000; Wallenstein, Eichenbaum, & Hasselmo, 1998; Alvarez & Squire, 1994) suggest that recollection of a recently encoded episode occurs when a retrieval cue activates a hippocampally guided representational pattern of cortical activity that encoded the episode. Through reciprocal hippocampocortical connections, this representation is reinstated in activation patterns in the cortex, allowing access to the initially encoded details of the episode. Consistent with

[1]RWTH Aachen University, Aachen, Germany, [2]University of Pennsylvania

this assumption, recollection of different classes of information is associated with activity in cortical regions that play a selective role in the on-line processing of the same information (e.g., Woodruff, Johnson, Uncapher, & Rugg, 2005; Wheeler & Buckner, 2004). Thus, cortical activity associated with successful recollection is at least partially dependent on the content of that memory (e.g., Woodruff et al., 2005).

Successful binding of elements within an event should also lead to a more accurate memory for the event. Areas activated during memory formation overlap with areas involved during retrieval in recognition (Weis, Klaver, Reul, Elger, & Fernandez, 2004). Using a ''subsequent memory procedure,'' numerous event-related fMRI studies have demonstrated that, relative to forgotten items, study items that are remembered elicit greater activity in a variety of cortical regions (e.g., Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998; Wagner et al., 1998; for a review see Paller & Wagner, 2002). Thus, several cortical processes during encoding are essential for subsequent memory performance and possibly demonstrate the multimodal features of an episode bound into a common memory representation.

In direct person-to-person communication, arm and hand gestures often accompany speech. Behavioral studies show that listeners use gestures in their comprehension of a speaker's overall message (Beattie & Shovelton, 1999, 2002; Goldin-Meadow, 1999; McNeill, 1999; Graham & Argyle, 1975) and that gestures can facilitate language production (e.g., Rose & Douglas, 2001; Hanlon, Brown, & Gerstman, 1990) and comprehension (Valenzeno et al., 2003; Kelly et al., 1999; Gauger, 1952). For example, Kelly et al. (1999) demonstrated that certain nonverbal behaviors, such as deictic and iconic gestures, can help people comprehend and remember pragmatic communication. Consistent with this finding, Valenzeno et al. (2003) showed that children learn when they watch a video containing the teacher's gestures as compared to videos without gestures. However, apart from the behavioral evidence for gesture related benefits in memory, the underlying neural processes are basically unknown.

In a previous fMRI study, the integration of speech and iconic gestures was investigated using a mismatch paradigm (Willems, Ozyurek, & Hagoort, 2007). The processing of mismatching information contained in action and speech differentially activated the left inferior frontal cortex (BA 45), the premotor cortex (BA 6), and the left STS. The authors concluded that activity in these regions was modulated by an increase in the semantic load resulting from conflicting speech and action information. The results of Willems et al. (2007) suggest that several cortical brain areas are involved in the processing of bimodal semantic information from speech and gesture, but the study did not report the effects of speech–gesture combinations on memory.

In behavioral studies, the processing of speech and corresponding gestures leads to a memory advantage (e.g., Valenzeno et al., 2003), even when the input modalities are different and associated at a high semantic level (e.g., Ozyurek, Willems, Kita, & Hagoort, 2007; Willems & Hagoort, 2007). For the formation of memory traces, integration and binding processes of these content dimensions are particularly relevant. We assume that this importance is increased when speech and gesture are connected at the level of abstract concepts. Metaphorical gestures are an important type of gestures that accompany speech. In contrast to iconic gestures they refer to an abstract rather than concrete content of speech (McNeill, 1992). For example, the sentence ''The presentation is on a very *high* level'' is accompanied by raising the right hand representing the ''high level'' of the presentation. Therefore, metaphoric gestures represent concrete features in order to illustrate abstract concepts contained in the utterance. In contrast, iconic gestures directly refer to concrete features of the world, as in the sentence ''The ball is *round*'' (round hand gesture) or ''The cottage is located on a very *high* mountain'' (lift of the right hand).

In the current study, we investigated the neural correlates of integration/binding processes during implicit encoding of speech and gestures and the effects of this binding on memory performance. We compared brain activation during watching of short video clips of (a) meaningful metaphoric gesture (MG), (b) unrelated free gesture (FG), and (c) no gesture (NG) in reference to abstract sentence content. To obtain memory-related activations, we employed multiple regression analyses for each video condition with the memory performance of our participants (hits and d'), acquired in a subsequent recognition paradigm. Behaviorally, we expected the best memory performance for the meaningful metaphoric gesture condition. On the neural level, we hypothesized that the individual subsequent memory performance is related to cortical (for the gesture conditions predominantly in the inferior frontal gyrus [IFG]; e.g., Willems et al., 2007) and hippocampal (e.g., Eichenbaum, 2004) activations during encoding. Furthermore, we assumed distinct activation patterns predominantly in parts of the left IFG (Willems et al., 2007) and left temporal areas (Holle, Gunter, Rüschemeyer, Hennenlotter, & Iacoboni, 2008) for related metaphoric gestures, indicating semantic integration.

METHODS

Participants

Twelve male right-handed (Oldfield, 1971) healthy volunteers, all native speakers of German (mean age = 28.96 years, range = 23–38 years), without impairments of vision or hearing, participated in the study. None of the participants had any medical, neurological, or psychiatric illness, past or present. All participants gave written informed consent, and were paid €20 for

participation. The study was approved by the local ethics committee.

Stimulus Construction

A set of 1296 (162 × 8) short video clips depicting an actor was initially created: (1) German sentences with abstract contents and corresponding metaphoric gestures; (2) German sentences with abstract contents and unrelated, free gestures; (3) Russian sentences and corresponding metaphoric gestures; (4) Russian sentences and unrelated, free gestures; (5) German sentences with abstract contents in isolation (without gestures); (6) Russian sentences in isolation (without gestures); (7) metaphoric gestures in isolation (without speech); and (8) unrelated, free gestures in isolation (without speech).

The current study focuses on three stimulus types: (1) German sentences with abstract contents and corresponding metaphoric gestures (MG); (2) German sentences with abstract contents and unrelated, free gestures (FG); and (3) German sentences with abstract contents in isolation (no gesture: NG). Metaphoric gestures constitute one of McNeill's (1992) basic gesture types and resemble something concrete in order to represent something abstract. For example, the sentence ''The twins had a spiritual *bond* between them'' is accompanied by a short connection of the fingertips of both hands in the middle of the body representing the ''bond'' between them.

All 1296 sentences have the same length of five to eight words, with an average duration of 2.47 sec (SD = 0.4) and a similar grammatical form (subject–predicate–object). The speech and gestures were performed by the same male actor in a natural, spontaneous way. In the MG condition, the actor illustrated the content of the sentences with semantically related metaphoric gestures. In the FG condition, we included gestures to the same sentences with similar complexity, performed by the same hand/hands and with movements in similar directions, however, with little or no semantic relation to the sentence context. In the NG condition, there was no gesture accompanying the same sentence. This procedure was continuously supervised by two of the authors (B. S. and A. G.) and timed digitally. All video clips have the same length of 5 sec with at least 0.5 sec before and after the sentence onset and offset, respectively, where the actor neither speaks nor moves.

For stimulus validation, 20 raters who did not take part in the fMRI experiment rated each video on a scale from 1 to 7 on understandability, imageability, and naturalness (1 = very low to 7 = very high). Other parameters, such as movement characteristics, pantomime content, transitivity, or handedness, were coded by two of the authors (B. S. and A. G.). A set of 1024 video clips (128 abstract sentences with metaphoric gestures and their counterparts in the other seven conditions) were chosen from the total pool of stimuli for the fMRI experiment on the basis of high naturalness and balanced movement characteristics, as well as high understandability for the German conditions. Out of these 1024 videos, four homogeneous sets were created such that each participant was presented with 256 sentences during the scanning procedure, counterbalanced across the subjects so that one subject did not see gesture or speech repetitions of any single item. In each of the four complementary sets, only those two conditions of an item were repeated which contain different speech and gesture information. For example, German sentences with abstract contents and corresponding metaphoric gestures (Condition 1) and Russian sentences and unrelated, free gestures (Condition 4). Further examples are Conditions 2 and 7, 3 and 5, or Conditions 6 and 8 (see above). Across all participants, each item occurred in each condition to control for possible differences in stimulus characteristics and sentence contents.

The semantic relatedness of speech and gesture was explicitly manipulated in the video conditions FG (low semantic relatedness) and MG (high semantic relatedness) of our experiment. To confirm the quality of our manipulation, in 12 healthy subjects we conducted additional ratings of the semantic relatedness of speech and gesture in the bimodal conditions. Our results support the validity of our manipulation: In all four stimulus sets, speech and gesture in the MG condition was shown to be semantically more related in comparison to the FG condition [Set 1—MG: M = 5.57, SD = 1.09; FG: M = 2.76, SD = 0.967; Set 2—MG: M = 5.85, SD = 0.50; FG: M = 2.66, SD = 1.07; Set 3—MG: M = 5.62, SD = 1.01; FG: M = 2.94, SD = 1.14; Set 4—MG: M = 5.71, SD = 0.76; FG: M = 2.86, SD = 1.20; $F(1, 62) = 143.569$, $p < .001$, $\eta^2 = .698$; condition effect, within-subject ANOVA].

There were no significant differences in the rating parameters (''understandability,'' ''imageability,'' and ''naturalness'') between the stimulus sets [main effect: set $F(9, 1860) = 0.447$, $p = .91$, $\eta^2 = .002$; interaction effect: Set × Condition $F(36, 1860) = 0.655$, $p = .94$, $\eta^2 = .013$; multivariate ANOVA]. Thus, the following statistics are based on all 128 items (4 × 32) for each condition together.

For both combined conditions (MG, FG), we found no significant differences in understandability in comparison to the NG condition (MG: M = 6.77, SD = 0.18; FG: M = 6.53, SD = 0.45; NG: M = 6.66, SD = 0.19; MG vs. NG: $p = .639$; FG vs. NG: $p = .311$). In contrast, direct comparisons between MG and FG indicated better comprehension in the MG condition [$t(254) = 5.59$, $p < .001$]. Despite this difference, video clips including German language scored higher than 6 on understandability [MG: M = 6.77, SD = 0.18, $t(127) = 48.86$, $p < .001$; FG: M = 6.53, SD = 0.45, $t(127) = 13.35$, $p < .001$; and NG: M = 6.66, SD = 0.19, $t(127) = 37.44$, $p < .001$; one-sample t test]. These results represent excellent

comprehension of the sentences with abstract contents independent of the gesture condition.

In contrast to the spoken sentences, gestures in isolation without a sentence context are relatively meaningless and were rated below 3.5 in understandability [isolated metaphoric gestures (IMG): $M = 2.94$, $SD = 0.71$, $t(127) = -8.89$, $p < .001$; isolated free gestures (IFG): $M = 2.40$, $SD = 0.58$, $t(127) = -21.69$, $p < .001$; one-sample t tests]. IMGs were rated higher than IFGs in understandability [$t(254) = 6.76$, $p < .001$].

Despite the same content in isolated and combined conditions, the MG condition was rated higher in imageability than its isolated counterparts (MG: $M = 4.52$, $SD = 0.52$; IMG: $M = 3.25$, $SD = 0.56$; NG: $M = 3.01$, $SD = 0.39$; MG > IMG: $p < .001$; MG > NG: $p < .001$). The same effect was found for the FG condition (FG: $M = 3.34$, $SD = 0.47$; IFG: $M = 2.85$, $SD = 0.47$; NG: $M = 3.01$, $SD = 0.39$; FG > IFG: $p < .001$; FG > NG: $p < .001$).

Similar to the results of the imageability, the MG condition was rated higher in naturalness than their isolated counterparts (MG: $M = 4.74$, $SD = 0.57$, IMG: $M = 3.90$, $SD = 0.58$; NG: $M = 3.95$, $SD = 0.33$; MG > IMG: $p < .001$; MG > NG: $p < .001$). In contrast, the FG condition was rated lower in naturalness than its isolated counterparts (FG: $M = 3.20$, $SD = 0.59$; IFG: $M = 3.39$, $SD = 0.47$; NG: $M = 3.95$, $SD = 0.33$; FG < IFG: $p < .05$; FG < NG: $p < .001$).

Direct comparisons of the combined conditions indicate that items of the metaphoric speech–gesture condition scored highest on all parameters [understandability—MG: $M = 6.77$, $SD = 0.18$; FG: $M = 6.53$, $SD = 0.44$; $t(254) = 5.59$, $p < .001$; imageability—MG: $M = 4.51$, $SD = 0.52$; FG: $M = 3.34$, $SD = 0.47$; $t(254) = 18.82$, $p < .001$; naturalness—MG: $M = 4.74$, $SD = 0.57$; FG: $M = 3.20$, $SD = 0.59$; $t(254) = 21.30$, $p < .001$]. These differences may be a direct effect of the manipulation of semantic relatedness between speech and gesture because the sentence content is the same for each condition and the evaluation parameters of the isolated speech or isolated gesture condition cannot explain the differences in the bimodal conditions.

Recorded sentences did not differ significantly in average duration measured from speech onset to speech offset [MG: $M = 2532$ msec, $SD = 382$ msec; FG: $M = 2489$ msec, $SD = 345$ msec; and NG: $M = 2470$ msec, $SD = 366$ msec; $F(2, 383) = 0.979$, $p = .377$; one-way ANOVA]. Gesture duration, measured from arm movement onset to arm movement offset, had the same length in both gesture conditions [MG: $M = 2394$ msec, $SD = 418$ msec and FG: $M = 2422$ msec, $SD = 365$ msec; $t(254) = -0.581$, $p = .561$].

Furthermore, for the MG condition, ''points of integration'' were defined as the time point of the highest connection between speech and gesture content. For example, for the sentence ''The twins had a spiritual *bond* between them,'' the integration point was set to the end of the word (''bond'') corresponding to the metaphoric gesture (connection of the fingertips). The time point of integration was transferred to the other conditions (FG and NG). This transfer was possible because of the standardized timing, form, and structure of each of the eight conditions corresponding to an item. These integration points occurred, on average, 2557 msec ($SD = 707$ msec) after the video start (2057 msec after speech onset).

Experimental Design and Procedure

During the fMRI scanning procedure, videos were presented via MR-compatible video goggles (stereoscopic display with up to 1024 × 768 pixel resolution; VisuaStim XGA, Resonance Technology) and nonmagnetic headphones (audio-presenting systems for stereophonic stimuli: Commander XG; Resonance Technology), which also dampened scanner noise.

Thirty-two items of each of the eight conditions were presented in an event-related design, in a pseudorandomized order and counterbalanced across subjects, so that one subject did not see gesture or speech repetitions of any single item. Across all participants, every video was presented in each of its eight conditions, but each participant saw only complementary items of the eight possible derivatives, thus, the same sentence or gesture information was never seen twice per participant. All videos had a duration of 5 sec and were followed by a baseline condition (gray background with a fixation cross) with a variable duration of 3750 to 6750 msec (average: 5000 msec).

Subjects were instructed to watch the videos and to respond at the beginning of each one by pressing one of two buttons with the left index and middle fingers to indicate whether a spot displayed on the actor's sweater was dark or light (see Figure 1 for a dark example). This implicit encoding task, without any memory instruction, was chosen to focus participants' attention on the middle of the screen without instruction biases using a manipulation independent of the video conditions. This enabled us to investigate implicit speech and gesture encoding. Before the scanning session, each participant received at least 10 practice trials outside the scanner, which were different from those used in the main experiment. To adjust the volume level of the video clips, during the overview scans additional clips were presented and the volume was adjusted. Each participant performed four runs with 64 video clips and a total duration of approximately 11 min each.

Behavioral Data Acquisition

Recognition memory performances for the three conditions (MG, FG, and NG) were examined about 10 to 15 min after scanning. All videos of the MG and FG conditions (32 each) and half of the NG condition (16)

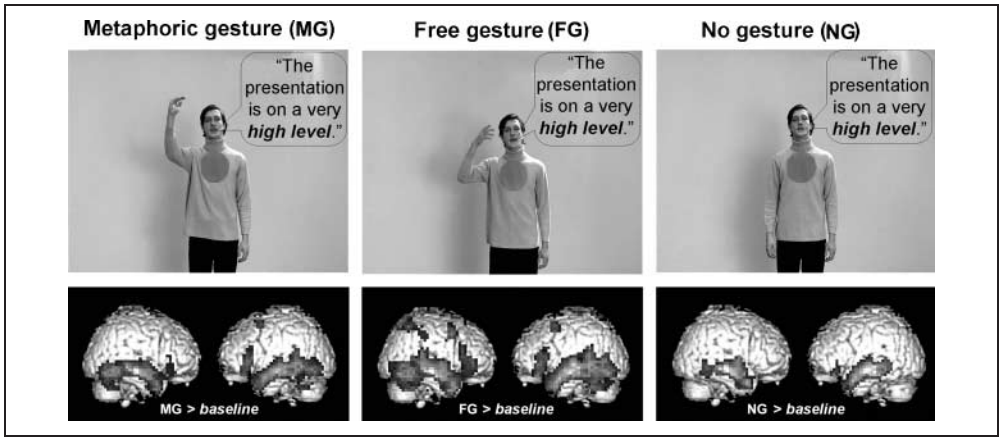

Figure 1. Examples of the different speech and gesture video clips and corresponding brain activation patterns against baseline. The stimulus material consisted of video clips of an actor performing meaningful metaphoric gestures (''MG''), unrelated, free gestures (''FG''), and no gesture (''NG'') together with abstract, metaphorical sentences. One screen shot of an example video is shown for each condition. The spoken German sentences are translated into English and written in the speech bubble of each picture for illustration (unlike in the actual stimuli). The activation patterns against baseline (fixation cross) for each condition are presented below each example.

were presented intermixed with an equal number of new items for each condition. Participants had to indicate via a ''yes''/''no'' response if they had seen the presented video before (''old'': left button) or not (''new'': right button). Altogether, 160 videos of the three conditions were presented randomly distributed in the recognition phase. Again, Presentation software (Version 9.2, 2005) was used for stimulus presentation and response measurement.

fMRI Data Acquisition

MRI was performed on a 3-T Philips scanner (Philips MRT Achieva series). Functional data were acquired with echo-planar images in 31 transversal slices (repetition time [TR] = 2000 msec; echo time [TE] = 30 msec; flip angle = 90; slice thickness = 3.5 mm; interslice gap = .35 mm; field of view [FoV] = 240 mm, voxel resolution = 3.5 × 3.5 mm). Slices were positioned to achieve whole brain coverage. During each functional run, 325 volumes were acquired. After the functional runs, for each participant, an anatomical scan was acquired using a high-resolution T1-weighted 3-D sequence consisting of 180 sagittal slices (TR = 9863 msec; TE = 4.59 msec; FoV = 256 mm; slice thickness = 1 mm; interslice gap = 1 mm).

Data Analysis

MR images were analyzed using Statistical Parametric Mapping (SPM2; www.fil.ion.ucl.ac.uk) implemented in MATLAB 6.5 (Mathworks, Sherborn, MA). The first 5 volumes of every functional run were discarded from the analysis to minimize T1-saturation effects. To correct for their different acquisition times, the signal measured in each slice was shifted relative to the acquisition time of the middle slice using a slice interpolation in time. All images of one session were realigned to the first image of a run to correct for head movement and normalized into standard stereotaxic anatomical MNI space by using the transformation matrix calculated from the first EPI scan of each subject and the EPI template. Afterwards, the normalized data with a resliced voxel size of 4 × 4 × 4 mm were smoothed with a 6-mm FWHM isotropic Gaussian kernel to accommodate intersubject variation in brain anatomy. Proportional scaling with high-pass filtering was used to eliminate confounding effects of differences in global activity within and between subjects.

The expected hemodynamic response at the defined ''points of integration'' for each event-type was modeled by two response functions, a canonical hemodynamic response function (Friston et al., 1998) and its temporal derivative. The temporal derivative was included in the model to account for the residual variance resulting from small temporal differences in the onset of the hemodynamic response, which is not explained by the canonical hemodynamic response function alone. The functions were convolved with the event sequence, with a fixed event duration of 1 sec for the onsets corresponding to the integration points of gesture stroke and sentence keyword to create the stimulus conditions in a general linear model. The fixed event duration of 1 sec was chosen to get a broader range of data around the assumed time point of integration because the time courses of speech and gesture integration processes in the brain are almost unknown.

A group analysis was performed for the main effects by entering contrast images into one-sample *t* tests, in which subjects are treated as random variables. Voxels

with a significance level of $p < .05$, corrected for the false discovery rate (FDR), belonging to clusters with at least 20 voxels are reported.

Multiple regression analyses without constant were performed on subject level. These analyses were chosen in contrast to analyses of individual stimuli because these analyses offer the opportunity to deal even with small numbers of events (e.g., hits, false alarms, correct rejections, misses) and take into account all events of a condition. Two factors were included for each regression analysis and condition: firstly, the hit and false alarm rate (FA), and secondly, the discrimination performance (d') and the response criterion (c). A Monte Carlo simulation of the brain volume of the current study was conducted to establish an appropriate voxel contiguity threshold (Slotnick, Moo, Segal, & Hart, 2003). Assuming an individual voxel type I error of $p < .001$, a cluster extent of 10 contiguous resampled voxels was indicated as necessary to correct for multiple voxel comparisons at $p < .01$. In reference to the relatively small sample size, the rather strict corrected significance level should ensure the validity of our results.

For the extraction and presentation of hippocampal activation, multiple regression analyses without constant were performed, thresholded at a significance level of $p < .05$, uncorrected, belonging to clusters with at least 37 voxels. At individual voxel type I error of $p < .05$, a cluster extent of 37 contiguous resampled voxels was indicated as necessary to correct for multiple voxel comparisons at $p < .01$. To identify hippocampal activations, ROIs were applied to these thresholded analyses, using the small volume correction implemented in SPM2. These ROIs were defined by the probability maps from Amunts et al. (2005) and include all defined subregions of the hippocampal formation (CA, EC, SUB, FD, HATA). The exact anatomical regions of the resulting activation peaks were further defined, using again the probability maps from Amunts et al., which are based on the cytoarchitectonic separation of hippocampal subregions from 10 human brains.

To investigate the direct relationships between the identified parameter estimates (beta values) of hippocampal activations of the three conditions and the individual subsequent memory performance (hits and d'), beta values were extracted with the VOI function from SPM2, using an 8-mm sphere for each activation peak. A sphere was used instead of the whole cluster activation to avoid possible differences due to dissimilar extensions of clusters in these analyses. The individual parameter estimates for each region were used as predictors in linear regression analyses, using the SPSS software (version 14.0 for Windows; SPSS, Chicago, IL, USA), to predict the proportion of hits, hits > FAs, and d' of our participants.

All reported voxel coordinates of activation peaks were transformed from MNI space to Talairach and Tournoux (1988) atlas space by nonlinear transformations (www.mrc-cbu.cam.ac.uk/Imaging/mnispace.html).

Statistical analyses of data other than fMRI were performed using SPSS version 14.0 for Windows (SPSS). The t tests, ANOVAs, and linear regression analyses were applied for the analyses. Greenhouse–Geisser correction was applied whenever necessary. Discrimination performance (d'), response criterion (c), and likelihood ratio (β) were calculated following the signal detection theory ($d' = z(\text{hits}) - z(\text{FA})$; $c = -1/2[z(\text{hits}) + z(\text{FA})]$; $\beta = -1/2 d'[z(\text{hits}) + z(\text{FA})]$; e.g., Macmillan & Creelman, 1991). Statistical analyses are two-tailed with α levels of significance of $p < .05$.

RESULTS

Behavioral Results

Implicit Encoding Task

The average reaction time for the implicit encoding task (''indicate the color of the spot on actor's sweater'') did not differ across colors [left button: $M = 1.04$ sec, $SD = 0.04$; right button: $M = 1.03$ sec, $SD = 0.36$; $t(11) = 0.642$, $p = .534$, paired t test] and conditions [MG: $M = 1.04$, $SD = 0.38$; FG: $M = 1.02$, $SD = 0.37$; NG: $M = 1.01$, $SD = 0.33$; $F(2, 22) = 1.063$, $p = .363$; within-subject ANOVA]. The participants performed with an average accuracy rate of 99.22% ($SD = 0.81$), which did not differ across the conditions [$F(2, 22) = 0.133$, $p = .877$; within-subject ANOVA].

Recognition Task

The average reaction time for the recognition task (''old/new decision'') did not differ across conditions [MG_old: $M = 3.91$, $SD = 0.27$; FG_old: $M = 3.96$, $SD = 0.25$; NG_old: $M = 3.94$, $SD = 0.24$; MG_new: $M = 3.97$, $SD = 0.23$; FG_new: $M = 3.84$, $SD = 0.28$; NG_new: $M = 3.94$, $SD = 0.24$; $F(5, 55) = 2.07$, $p = .134$, within-subject ANOVA].

Analysis of Hits and False Alarms

The proportion of study items correctly endorsed as old (hits) was 60% for MG, 49% for FG, and 42% for the NG condition, with an FA rate of 34%, 19%, and 28%, respectively. A 2 × 3 within-subject ANOVA, with the factors memory performance (hits and FA) and gesture condition (MG, FG, and NG), revealed significant main effects for memory performance [$F(1, 11) = 26.30$, $p < .001$, $\eta^2 = .705$] and gesture condition [$F(2, 10) = 6.721$, $p < .05$, $\eta^2 = .379$], and a significant interaction of both factors [$F(2, 10) = 5.95$, $p < .05$, $\eta^2 = .543$]. Pairwise contrasts against the NG condition revealed that the interaction effect with memory (hits vs. FA) resulted of significant differences between FG and NG [$F(1, 11) = 12.89$, $p < .005$, $\eta^2 = .540$] and marginal differences between MG and NG [$F(1, 11) = 4.34$, $p = .061$, $\eta^2 = .283$]. Additional pairwise comparisons for hits and FA,

respectively, indicated that the hit rate was significantly enhanced in MG in comparison to FG [$t(11) = 3.60$, $p < .005$] and NG [$t(11) = 2.80$, $p < .05$; see Figure 2], and that the FA rate is significantly reduced in FG in comparison to MG [$t(11) = -4.97$, $p < .001$] and NG [$t(11) = 2.36$, $p < .05$; see Figure 2].

Despite the fact that the hit rate is above chance (50% correct) only for the metaphoric gesture condition [MG: $t(11) = 2.98$, $p = .013$; FG: $t(11) = -0.14$, $p = .889$; NG: $t(11) = -1.36$, $p = .203$; see Figure 2], the guess corrected recognition performance (hits > FA) is different from zero for all conditions [MG: $t(11) = 4.36$, $p < .001$; FG: $t(11) = 7.14$, $p < .001$; NG: $t(11) = 2.57$, $p < .05$; see Figure 2]. Furthermore, as indicated by the contrasts of the ANOVA, with one-tailed t tests, we obtained significant gesture-related memory advantages (as hits > FA) in comparison to the NG condition for the FG [$t(11) = 3.60$, $p < .003$] and the MG conditions [$t(11) = 2.09$, $p = .031$, one-tailed t tests]. There were no differences between the gesture conditions [MG vs. FG: $t(11) = -1.16$, $p = .274$; two-tailed t tests]. Correlation analyses indicated that there was no significant relation between hits and FA for all conditions (MG: $r = .063$, $p = .847$; FG: $r = .166$, $p = .605$; NG: $r = .418$, $p = .176$).

To control for the impact of meaningful gestures alone to the subsequent memory performance for the bimodal conditions, we calculated the recognition probability of any single item in the bimodal conditions and further correlated them with the understandability scores of the corresponding isolated gestures. The recognition probability of any single item was calculated as the average number of subjects that correctly remembered the item, divided by the total number of participants that saw the item. There was no relation between understandability of the gestures in isolation (IMG and IFG together) and the recognition performance of their bimodal counterparts (MG and FG together; Set 1: $r = .072$, $p = .570$; Set 2: $r = .066$, $p = .607$; Set 3: $r = .138$, $p = .276$; Set 4: $r = .126$, $p = .320$).

Analysis of Discrimination Performance (d') and Response Criteria (c)

The analysis of hits and FA indicated differences in response behavior of the participants across the conditions. Especially for the FG condition, participants showed the lowest FA rate across the three conditions but also fewer hits in contrast to the MG condition (see above). To identify the discrimination performance (d') between old and new items independent of the individual response criteria (c), a signal detection analysis (e.g., Macmillan & Creelman, 1991) was applied to all conditions.

For MG, the average of the old/new discrimination performance (d') was 0.681 ($SD = 0.536$) and significantly different from zero [$t(11) = 4.40$, $p < .001$]. In contrast, the response criteria (c) and the corresponding likelihood ratio (β) were not significantly different from zero [c: $M = 0.079$, $SD = 0.291$, $t(11) = 0.94$, $p = .366$; β: $M = 0.110$, $SD = 0.213$, $t(11) = 1.79$, $p = .101$], indicating an unbiased response behavior.

For FG, the average discrimination performance (d') was 0.879 ($SD = 0.430$) and significantly different from zero [$t(11) = 7.08$, $p < .001$]. However, the response criteria (c) and the corresponding β value were also significantly different from zero [c: $M = 0.461$, $SD = 0.287$, $t(11) = 5.56$, $p < .001$; β: $M = 0.422$, $SD = 0.309$; $t(11) = 4.73$, $p < .001$], indicating a conservative response behavior (tendency to press ''NO'').

For NG, the average of the discrimination performance (d') was 0.450 ($SD = 0.610$) and significantly different from zero [$t(11) = 2.55$, $p < .05$]. Despite the fact that the response criteria (c) were significantly different from zero [c: $M = 0.458$, $SD = 0.513$; $t(11) = 3.09$, $p < .05$], the corresponding β value is not [β: $M = 0.234$, $SD = 0.612$; $t(11) = 1.33$, $p = .212$], indicating a relatively unbiased response behavior.

The discrimination performance (d') for the MG condition was not significantly different from the FG [$t(11) = -1.93$, $p < .08$] and NG conditions [$t(11) = 1.31$, $p < .22$]. In contrast, response criteria (c) were significantly

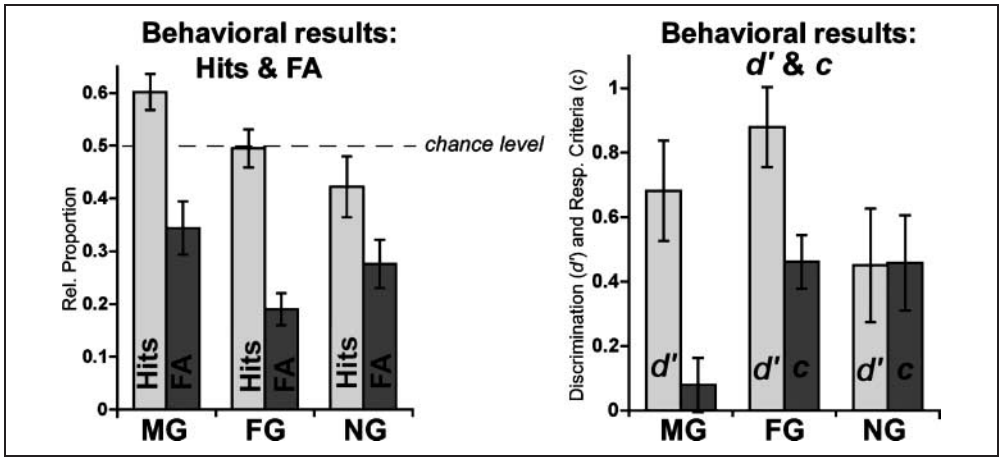

Figure 2. Behavioral results of hits and false alarms (FA), as well as discrimination performance (d') and response criteria (c), for the MG, FG, and NG conditions. This figure shows the behavioral memory performances of our participants as hits and false alarms (FA; left), and as discrimination performance (d') and response criteria (c) from a signal detection analysis (right). (MG = metaphoric gesture condition [red]; FG = free gesture condition [green]; NG = no-gesture condition [blue]). The error bars indicate the standard error of the mean.

increased in the FG [$t(11) = -6.20$, $p < .001$] and NG conditions [$t(11) = -2.32$, $p < .05$] in contrast to the MG condition. Only the β values of the FG condition were significantly increased in comparison to the MG condition [FG > MG: $t(11) = 3.68$, $p < .005$; MG vs. NG: $t(11) = -0.68$, $p = .512$].

The FG condition, in contrast to the NG condition, showed an increase in d' [$t(11) = -2.95$, $p < .05$] and no difference in the response criteria [$t(11) = 0.02$, $p < .982$] and β value [$t(11) = 1.04$, $p < .322$].

These results suggest different response criteria for the MG in contrast to the FG and NG conditions. Specifically, in the FG condition, the participants showed a conservative response behavior (tendency toward a "NO" response) which is significantly different from the MG condition. In consideration of these differences in response behavior between the conditions, the discrimination performance (d') indicates no increase in memory performance of the MG in contrast to the FG or NG condition. Only the FG condition showed an increased discrimination performance (d') in contrast to the NG condition.

fMRI Results

Main Effects

For all three conditions (MG, FG, NG) versus baseline (fixation cross), we found an extended network of bihemispheric medial and lateral temporal and left inferior frontal activations. In both gesture conditions (MG and FG), activation clusters extended into the occipital lobes, the cerebellum, as well as into parietal and motor areas (see Figure 1 and Table 1).

Table 1. Activation Peaks and Cluster Extensions of the MG, FG, and NG Conditions against Baseline (Fixation Cross)

				Coordinates				
Anatomical Region	*Cluster Extent*	*Hemisphere*	*BA*	*x*	*y*	*z*	*t*	*No. Voxels*
Metaphoric Coverbal Gestures (MG)								
Superior temporal gyrus	MTG, IFG	R	42	59	−19	8	14.96	603
Middle temporal gyrus	STG, IFG	L	21	−51	3	−17	9.60	720
Premotor cortex		L	6	−40	11	55	6.98	24
Postcentral gyrus		R	4, 5	4	−39	68	6.07	31
Premotor cortex			6	0	11	66	5.31	27
Hippocampus		R	28	16	−12	−16	5.31	27
Hippocampus	MTL	L	34	−20	−9	−16	5.08	67
Free Coverbal Gestures (FG)								
Middle temporal gyrus	bilateral TL, OL, IFG, and MTL (incl. HC)	R	21	59	−20	−6	13.08	2648
Postcentral gyrus	Parietal lobe	R	4	8	−39	72	7.64	298
Medial frontal cortex	Premotor cortex	R	8	51	10	47	6.92	52
Premotor cortex		L	6	−44	10	51	6.66	36
Premotor cortex		R	6	8	7	66	6.29	55
Caudate nucleus		R		20	−18	19	6.17	31
No Gestures (NG)								
Middle temporal gyrus	STG	R	21	63	−24	−6	9.34	392
Middle temporal gyrus	STG, IFG	L	21	−55	−20	−6	9.01	408
Hippocampus	MTL	L	28	−20	−16	−16	6.14	61

Significance level (t value) and size of the respective activation cluster (number of voxels > 20) at $p < .05$, FDR corrected. Coordinates are listed in Talairach and Tournoux (1988) atlas space. BA is the Brodmann's area nearest to the coordinate and should be considered approximate. Cluster extensions denominate activated regions for larger voxel clusters encompassing different brain areas and should be considered approximate.

OL = occipital lobe; TL = temporal lobe; IFG = inferior frontal gyrus; STG = superior temporal gyrus; MTL = medial temporal lobe; HC = hippocampus.

Multiple Regressions of Hits and Hits > False Alarms (Hits > FA)

To obtain activations related to successful encoding, we calculated separate multiple regression analyses for each condition (MG, FG and NG), including both one factor for the individual hit rate and one for the FA rate. Both factors were chosen to independently represent the correct and false response proportions in the subsequent recognition task with the purpose to calculate the guess corrected recognition performance (hits > FA). The typical hits > misses contrast for subsequent memory performance is not appropriate for analyses across subjects because hits and misses (as well as FA and correct rejections [CR]) are negatively correlated (with $r = -1$) and explain the same amount of variance. Results are shown for each condition in Figure 3 (MG: red; FG: green; NG: blue) and Table 2, with positive correlations of activation pattern of hits and the difference contrast of hits > FA (contrast weights 1 [hits] and −1 [FA]). The latter contrast represents brain activations specifically related to memory performance (hits) and not to failures (FA).

For the MG condition, a positive correlation with the number of hits with bilateral activations of the IFG and the MTG was found (see Table 2). For the difference contrast, hits > FA, we obtained a left-lateralized activation of the IFG (see Figure 3 [red and yellow] and Table 2).

For the FG condition, we obtained a positive correlation for the hits with bilateral activations of the occipito-temporal–cerebellar region and the IFG (see Table 2). For the difference contrast, hits > FA, we obtained a left lateralized activation of the IFG and the occipito-temporal–cerebellar region (see Figure 3 [green] and Table 2).

For the NG condition, there was no activation for the correlation with the subsequent hit or hit > FA rate present.

Taken together, our results indicate that predominantly inferior frontal regions are differentially involved in memory processes related to the hit rate for speech and gesture videos (MG and FG). For both gesture conditions, we obtained activations of parts of the left IFG (BA 45, $tal_{x, y, z}$: −55, −27, −6; $t = 6.17$; conjunction analyses, $p < .01$, corrected) in relation to the hit rate.

Multiple Regressions of Discrimination Performance (d') and Response Criteria (c)

As indicated by the analysis of the behavioral data (see above), the brain activation in relation to the raw proportions of hits and FA may be affected by differences in response biases of the participants across the conditions. To obtain activations related to successful encoding independent of differences in the response criteria, we calculated separate multiple regression analyses for each condition (MG, FG, and NG), including both one factor for the individual discrimination performance (d') and one for the response criteria (c). Both factors were chosen to represent independently the encoding performance (represented in d') and the response bias (c) in the recognition task. Results are shown in Figure 4A and Table 3, with positive correlations of activation pattern with d', for each condition (MG: red; FG: green; NG: blue).

For the MG condition, a positive correlation of the discrimination performance (d') with activations of the left IFG (BA 45/47), the premotor cortex (BA 6), and the left MTG (BA 21), as well as an additional right-hemispheric activation of the IFG (BA 47; see Table 3), was revealed.

For the FG condition, we revealed a positive correlation of d' with activations located in the left occipito-temporal–cerebellar region and the right IFG (BA 47; see Table 3).

For the NG condition, we also revealed a positive correlation of d' with activations in the left occipito-temporal–cerebellar region (see Table 3). With regard to the response criteria (c), no activation reached the significance level.

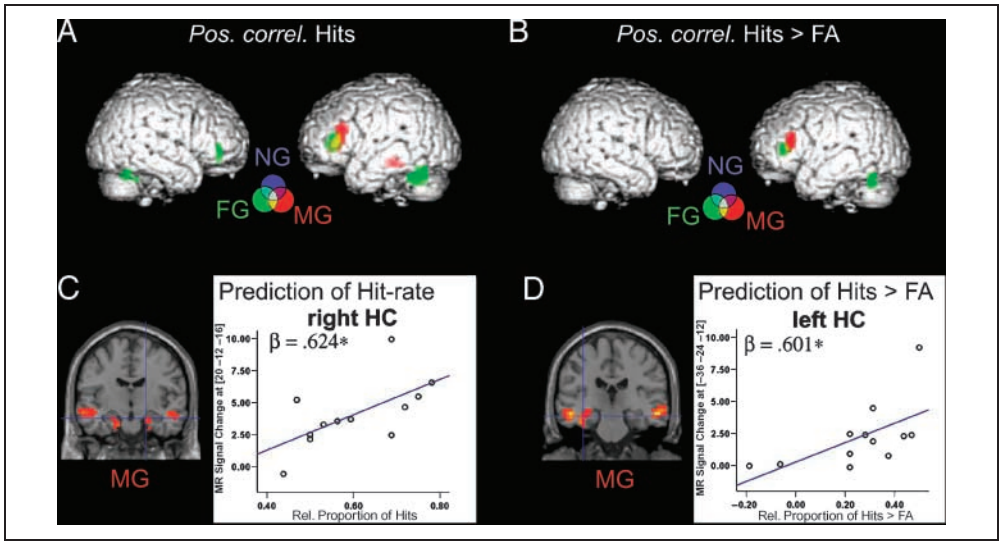

Figure 3. Neural correlates of hits and hits > false alarms (hits > FA) for the three gesture conditions. A and B show the correlations of activation patterns with the individual subsequent hits (A) and hits > FA (B) for each condition, respectively (MG = red; FG = green; NG = blue; MG ∩ FG = yellow). C and D depict the hippocampal activations in relation to subsequent hits (C) and hits > FA (D; see Table 4). The plots on the right side next to each picture show the results of linear regression analyses of the MR signal change (arbitrary units) and the hits and hits > FA (see Methods).

Table 2. Positive Correlations for Hits and Hits > False Alarms (Hits > FA) were Obtained for the Metaphoric Gesture (MG), the Free Gesture (FG), and No Gesture (NG) Conditions

				Coordinates				
Contrast	*Anatomical Region*	*Hemisphere*	*BA*	*x*	*y*	*z*	*t*	*No. Voxels*
Metaphoric Gesture Condition (MG)								
Positively correlated hits	Inferior frontal gyrus	L	45/47	−51	28	10	8.30	61
	Middle/inferior temporal gyrus	L	21/37	−51	−28	−9	5.76	18
Positively correlated hits > FA	Inferior frontal gyrus	L	45	−51	28	10	6.42	26
Free Gesture Condition (FG)								
Positively correlated hits	Occipito-temporal–cerebellar region	L	37/19	−44	−63	−20	10.58	49
		L	37/19	−44	−52	−24	6.51	
	Inferior frontal gyrus	R	47	51	42	−5	9.44	10
	Inferior frontal gyrus	L	47	−55	31	6	7.19	36
	Occipito-temporal–cerebellar region	R	37/19	48	−59	−17	5.56	11
		R	37/19	44	−59	−24	4.88	
		R	37/19	48	−44	−21	4.75	
Positively correlated hits > FA	Occipito-temporal–cerebellar region	L	37/19	−44	−63	−20	7.91	18
	Inferior frontal gyrus	L	45	−55	31	6	5.43	12

Significance level (t values) and size of the respective activation cluster (number of voxels) at $p < .01$, corrected. Coordinates are listed in Talairach and Tournoux (1988) atlas space. BA is the Brodmann's area nearest to the coordinate and should be considered approximate.

These results show different activation patterns in relation to d' for the three conditions, indicating gesture-related processing differences during encoding. Whereas predominantly language-related areas of the left hemisphere were activated for the MG condition in relation to the discrimination performance, the FG and NG conditions showed an involvement of the left occipito-temporal–cerebellar region (both) and the right IFG (only FG).

Hippocampal Activation during Encoding as Predictor of the Subsequent Memory Performance

We hypothesized that hippocampal involvement in speech and gesture binding leads to successful formation of a memory representation. To test this hypothesis, ROIs were applied to the multiple regression analysis for the subsequent hit rate and discrimination performances (see Data Analysis).

For the hit rate of the MG condition, we found three clusters within the ROI with corresponding activation peaks located in the left cornu amonis (CA, 80%), the left entorhinal cortex (EC, 100%), and the right subliminar cortex (SUB, 60%; see Table 4). The percentages are based on a sample of 10 human brains and indicate the probability for the identified activation peaks to correspond to the denominated anatomical subregion of the hippocampus (Amunts et al., 2005). For the discrimination performance (d') of the MG condition, the activation peaks were located in the left and right EC (90%, respectively) with an additional subpeak of the left HC in the CA (80%; see Table 4).

For the hit rate of the FG condition, the activation peaks were located in the right CA (40%) and right EC (50%), as well as two clusters in the left SUB cortex (30% and 70%; see Table 4). For the discrimination performance (d') of the FG condition, the activation peaks were located in the left EC (70%), the left SUB (50%), and the right EC (40%; see Table 4).

The peak activation for the hit rate of the NG condition was located in the left EC (100%; see Table 4). For the discrimination performance (d'), no activation occurred in the NG condition.

To investigate the specific relationship of the identified hippocampal activations with the subsequent memory performance, extracted activation patterns from each region were used as predictors for the hits, hits > FA, and d' in linear regression analyses.

The activation pattern extracted from the right SUB cortex activated during the MG condition showed a significant positive linear relationship ($\beta = .624$, $t = 2.525$, $p < .05$) with the subsequent hit rate (see Figure 3C and Table 4). In contrast, the activation of the left CA showed a positive linear relationship to the hits > FA rate ($\beta = .601$, $t = 2.381$, $p < .05$; Figure 3D and Table 4). The activation pattern extracted from the left CA for analysis of the discrimination performance (d') during the MG

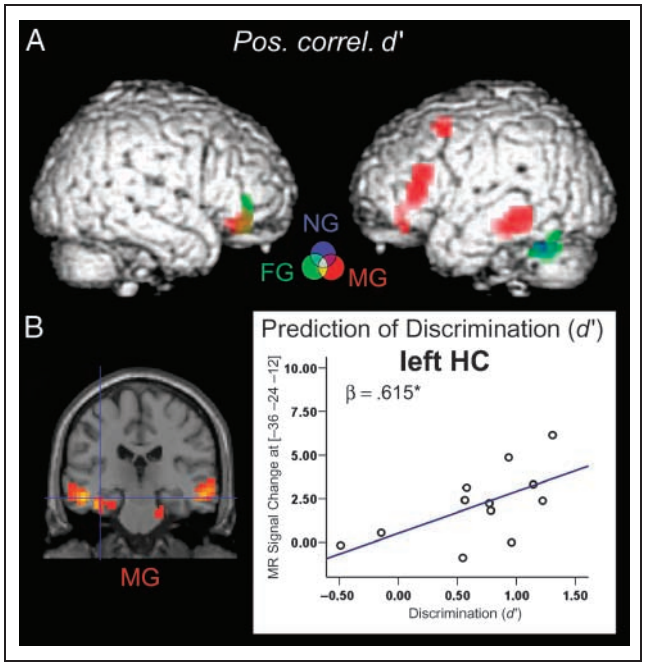

Figure 4. Neural correlates of discrimination performance (d') for the three gesture conditions. (A) The correlations of activation patterns with separate multiple regression analyses for the three conditions, including one factor for the individual discrimination performance (d') and another one for the subsequent response criteria (c), for each condition respectively (MG = red; FG = green; NG = blue; MG ∩ FG = yellow; FG ∩ NG = dark green). (B) The hippocampal activations in relation to subsequent d' (see Table 4). The diagram shows the result of the linear regression analysis of the extracted MR signal change (arbitrary units) and d' (see Methods).

condition showed a significant positive linear relationship with the subsequent discrimination performance ($\beta = .615$, $t = 2.463$, $p < .05$; see Figure 4B and Table 4) and hit > FA rate ($\beta = .580$, $t = 2.251$, $p < .05$; see Table 4). For the FG and NG conditions, no regression analysis reached significance level (see Table 4).

DISCUSSION

Gestures are a substantial component of direct human communication. In our study, we investigated the neural activity related to subsequent memory performance for videotaped spoken sentences which differed in the accompanying gestures: metaphoric coverbal gestures (MG), unrelated free coverbal gestures (FG), and no gestures (NG). In line with our assumptions, we found specific cortical and hippocampal activations related to the subsequent memory performance especially for the metaphoric coverbal gestures.

An analysis of the raw proportion of hits indicated an overlapping region for the MG and FG conditions located in the IFG in relation to the subsequent memory performance. In contrast, regression analyses of discrimination performances (d') independent of the response criteria (c) indicated different neural responses across conditions. Whereas the discrimination performance of the NG and FG conditions relied on left occipito-temporal–cerebellar regions (for NG and FG) and the right IFG (for FG), the encoding of the metaphoric condition

Table 3. Positive Correlations for Discrimination Performance (d') were Obtained for the Metaphoric Gesture (MG), the Free Gesture (FG), and No Gesture (NG) Conditions

				Coordinates				
	Anatomical Region	*Hemisphere*	*BA*	*x*	*y*	*z*	*t*	*No. Voxels*
Metaphoric Gesture Condition (MG)								
Positively correlated d'	Inferior frontal gyrus	L	45	−51	28	10	8.24	67
	Inferior frontal gyrus	L	47	−48	34	−15	4.81	
	Inferior frontal gyrus	L	47	−48	38	−9	4.52	
	Premotor cortex	L	6	−48	14	47	7.24	14
	Inferior frontal gyrus	R	47	51	38	−12	7.07	17
	Middle temporal gyrus	L	21	−55	−43	−5	6.70	92
	Middle temporal gyrus	L	21	−55	−32	−12	6.63	
Free Gesture Condition (FG)								
Positively correlated d'	Inferior frontal gyrus	R	47	51	42	−5	6.60	13
	Occipito-temporal–cerebellar region	L	37/19	−48	−67	−13	6.59	24
No Gesture Condition (NG)								
Positively correlated d'	Occipito-temporal–cerebellar region	L	37/19	−51	−57	25	5.70	10

Significance level (t values) and size of the respective activation cluster (number of voxels) at $p < .01$, corrected. Coordinates are listed in Talairach and Tournoux (1988) atlas space. BA is the Brodmann's area nearest to the coordinate and should be considered approximate.

Table 4. Activation Peaks and Linear Relationships with Behavioral Data for Activation in the Hippocampal Formation (CA, EC, FD, SUB, and HATA) Based on the Multiple Regression Analyses with Hit Rate and Discrimination Performance (d')

				Coordinates					*Beta Coefficients For*		
Condition	*Anatomical Region*	*Hemisphere*	*Prob.*	*x*	*y*	*z*	*t*	*No. Voxels*	*Hits*	*Hits > FA*	d'
MG: Hits	SUB	R	60%	20	−12	−16	3.02	20	*.624**	*.403*	*.373*
	EC	L	100%	−16	−9	−23	3.41	25	*.240*	*.365*	*.359*
	CA	L	80%	−36	−24	−12	2.99	17	*.114*	*.601**	*.620**
MG: d'	EC	R	90%	20	−9	−20	5.13	29	*.466*	*.353*	*.327*
	EC	L	90%	−16	−9	−23	4.16	59	*.272*	*.362*	*.356*
	EC	L	100%	−20	−5	−27	3.66		*.144*	*.412*	*.417*
	CA	L	80%	−36	−24	−12	2.38		*.229*	*.580**	*.615**
FG: Hits	CA	R	40%	40	−20	−12	3.95	11	*.013*	*.165*	*.280*
	EC	R	50%	28	2	−30	2.91	9	*.038*	*.143*	*.246*
	SUB	L	30%	−12	−35	−2	3.53	29	*.170*	*.494*	*.559*
	SUB/EC	L	70/60%	−16	−13	−20	3.50	20	*.295*	*.517*	*.545*
FG: d'	EC	L	70%	−16	−13	−20	2.88	15	*.271*	*.415*	*.461*
	SUB	L	50%	−16	−39	6	2.82	17	*.371*	*.532*	*.483*
	EC	R	40%	24	3	−20	2.59	12	*.394*	*.360*	*.305*
NG: Hits	EC	L	100%	−28	−13	−30	2.49	19	*.109*	*−.021*	*.463*

Significance level (t value) and size of the respective activation cluster (number of voxels) in the ROI of the hippocampus at $p < .01$, corrected for whole-brain analysis. The ROIs includes all hippocampal regions (CA, EC, SUB, FD, HATA) of the probability maps defined by Amunts et al. (2005) and the percentages in column ''Prob.'' refer to the precise probability of the location (anatomical region) of the activation peaks. Coordinates are listed in Talairach and Tournoux (1988) atlas space. Significant beta values (*$p < .05$) from regression analyses of extracted activation patterns (VOI function and 8 mm sphere) and hit and d' rate.

CA = cornu amonis; EC = entorhinal cortex; SUB = subiculum.

in relation to d' activated a left-lateralized network of the IFG, the BA 6 region, and the MTG, as well as the right IFG (BA 47). In addition, extracted hippocampal activation patterns were significantly related to the discrimination performance (d') and hit proportion only for the MG condition.

In summary, our results suggest that the effects of semantic integration or binding of speech and gesture on memory have a different neural instantiation than when such integration does not occur. The results are in line with studies showing IFG activation for semantic speech and gesture integration (e.g., Willems et al., 2007) and hippocampal involvement in multimodal memory binding processes (e.g., Eichenbaum, 2004).

Behaviorally, we found memory advantages as an increased hit rate for the MG condition in comparison to the hit rates of the FG and NG conditions. However, these differences can be explained by a more conservative response (as indicated by c) for the FG and NG conditions because we did not find significant differences in d' between MG and FG and between MG and NG. Despite the comparable discrimination performance across conditions, we obtained distinct memory-related left-hemispheric activations in the IFG, the premotor cortex (BA 6), and the MTG, as well as significant correlations between hippocampal areas and memory performance in the MG condition. These results are in line with other encoding studies which have shown that successful memory formation, measured as the difference in brain activity between subsequently remembered and forgotten items, is accompanied by activity increases in medial temporal and inferior prefrontal areas (e.g., Morcom, Good, Frackowiak, & Rugg, 2003; Casasanto et al., 2002; Strange, Otten, Josephs, Rugg, & Dolan, 2002; Davachi, Maril, & Wagner, 2001; Otten, Henson, & Rugg, 2001; Otten & Rugg, 2001; Kirchhoff, Wagner, Maril, & Stern, 2000; Brewer et al., 1998; Wagner et al., 1998; for a review, see Paller & Wagner, 2002). Furthermore, several fMRI studies support the involvement of the IFG in memory encoding for semantic tasks with verbal material (e.g., Henson, Hornberger, & Rugg, 2005; Fletcher, Stephenson, Carpenter, Donovan, & Bullmorel, 2003; Baker, Sanders, Maccotta, & Buckner, 2001; Davachi et al., 2001; Otten et al., 2001; Wagner et al., 1998), in particular, for items subsequently recollected (Davachi, Mitchell, & Wagner, 2003; Ranganath & Rainer, 2003; Henson, Rugg, Shallice, Josephs, & Dolan, 1999). This specific left-lateralized activation pattern

possibly reflects better semantic integration (binding) of speech and gesture information in our MG condition. In contrast, unrelated speech and gesture information seems to be processed similarly to the NG condition in left occipito-temporal (including BA 19/37) and cerebellar regions. Activations of the left cerebellum are in line with findings of Weis et al. (2004). They found left cerebellar activity to be related both to encoding and retrieval success using landscape images, possibly indicating a more visual–perceptual level of encoding. In contrast, Fliessbach, Trautner, Quesada, Elger, and Weber (2007) showed a common involvement of the left lateral cerebellum for collapsed semantic and nonsemantic encoding tasks. Kirchhoff and Buckner (2006) investigated the functional–anatomic correlates of individual differences in memory for picture associations. During intentional long-term memory encoding, they found occipito-temporal regions in relation to a visual inspection strategy (BA 37/19) and inferior frontal regions in relation to verbal elaboration (Kirchhoff & Buckner, 2006). Despite the fact that our participants were not instructed to memorize the video clips, the data of Kirchhoff and Buckner are consistent with the interpretation of a perceptual level of encoding in the FG and NG conditions and a more semantic encoding in the MG condition. The study of Prince, Daselaar, and Cabeza (2005) also indicates an involvement of the left lateral prefrontal cortex in semantic encoding processes and an involvement of left occipito-temporal regions (BA 19/37) during perceptual encoding. In a more recent study, Prince, Tsukiura, and Cabeza (2007) directly examined the effect of semantic relatedness of word pairs on the subsequent memory performance. They even found distinct parts of the IFG in relation to semantic relatedness and episodic memory performance: A posterior region was activated when the semantic link between word pairs was high, a mid region was involved in both episodic encoding and high semantic link, and an anterior region was involved in episodic encoding but only when the semantic link between word pairs was high (Prince et al., 2007). Although we could not differentiate between these IFG regions, we found a large activation cluster including BA 45 and BA 47 for high semantic relation (MG) correlating with the memory performance. From these results, at least the more anterior region (BA 47) is consistent with the findings from Prince et al. (2007). Together, these data support our interpretation of a semantic encoding in the MG condition and a rather perceptual level of encoding in the FG and NG conditions.

Besides these cortical activations, we found hippocampal activation in relation to subsequent memory performance. Consistent with this finding, it is generally agreed that old/new recognition is dependent on the integrity of the medial temporal lobes (MTLs), which include the hippocampal formation. In support of this idea, functional neuroimaging studies have directly related MTL activity to successful recognition performance (Daselaar et al., 2001; Donaldson, Petersen, & Buckner, 2001; Eldridge, Knowlton, Furmanski, Bookheimer, & Engel, 2000; Nyberg, Mcintosh, Houle, Nilsson, & Tulving, 1996) or reported subsequent memory effects in the MTL using yes/no recognition after semantic study tasks (Morcom et al., 2003; Otten et al., 2001; Wagner et al., 1998). We found activation of the right anterior MTL to be related to the subsequent hit rate. This finding is consistent with previous evidence that subsequent memory effects in anterior MTL reflect encoding of information that supports later recognition (Gold et al., 2006; Uncapher & Rugg, 2005; Ranganath & Rainer, 2003; Davachi & Wagner, 2002). Relational memory processes (e.g., Eichenbaum & Cohen, 2001; Cohen & Eichenbaum, 1993) are especially important in making semantic associations and have been linked to specific parts of the MTLs. Davachi and Wagner (2002), for example, found that the hippocampus was more active during relational (subjects incidentally encoded word triplets by ordering the words in terms of ''desirability'') than item-based processing (incidental encoding by rote repetition), which was associated with greater activity in entorhinal and parahippocampal regions. These data suggest that semantically related and unrelated information is encoded differently in distinct parts of the MTL. This functional differentiation of MTL structures with regard to relational memory is somewhat inconsistent with the observed entorhinal activation across all three conditions (MG, FG, and NG) in our study. However, the entorhinal cortex provides the majority of the input to the hippocampus (e.g., Witter & Amaral, 1991) and might elicit significant processing whether or not the hippocampus ends up being significantly engaged. The observed entorhinal activation across all three conditions might call into question our interpretation of binding of bimodal speech and gesture information. Although the activation of the EC across all conditions reflects encoding of a single modality (single-item encoding), we observed additional activation in the CA region of the hippocampus for the bimodal conditions. This is consistent with the results of Prince et al. (2007) showing a main effect of memory performance (episodic encoding) independent of semantic relatedness of word pairs in the left hippocampus. Furthermore, in the MG condition, we observed the strongest relation between the left CA region and the subsequent discrimination performance. This would indicate that, in addition to unimodal processes (within the entorhinal cortex), binding processes (within the hippocampus [CA]) during encoding predict subsequent memory performance (d'), especially in the MG condition. Thus, the components of the speech–gesture events that were processed and represented in different cortical regions possibly had been bound by the hippocampus into a common memory representation. The reactivation of this bimodal representation was necessary to give a correct response in the subsequent recognition task because participants were required to decide if they had seen the video including speech and gesture before.

Despite the consistent findings of MTL involvement in subsequent memory tasks across studies, there are conflicting results about the exact subregions of the hippocampal formation related to subsequent memory performance. Nevertheless, our results indicate a specific involvement of the hippocampus in the encoding processes of abstract spoken sentences accompanied by metaphoric gestures and possibly reflect the successful binding of the cortical representation of speech–gesture information (e.g., Norman & O'Reilly, 2003; Shastri, 2002; Rolls, 2000; Wallenstein et al., 1998; Alvarez & Squire, 1994). For the MG condition, subsequent discrimination performance (d') was specifically correlated with left inferior frontal, premotor, middle temporal, and hippocampal activations. These findings show that participants' behavior reflects the combined effects of multiple brain regions, possibly indicating a multifeature-binding substrate. For metaphoric gestures and speech, this effect was predominantly located in a left lateral language network of inferior frontal and temporal brain regions, indicating semantic integration (e.g., Willems et al., 2007). The additional activation of the left premotor cortex (BA 6) possibly represents the motor component of coverbal gesture processing and is in line with studies showing activation of the premotor cortex in response to arm and hand actions (Calvo-Merino, Glaser, & Grezes, 2005; Buccino et al., 2001; Hari et al., 1998). These differences in cortical involvement may reflect semantic integration of speech and gesture that facilitates speech comprehension.

Gestures are an important component in human face-to-face interaction and can play a number of functions in communication (see Goldin-Meadow, 1999; McNeill, 1992). The few investigations of speech and gesture binding (see Willems & Hagoort, 2007 for a review) have not examined this binding in relation to subsequent memory performance. We suggest that our findings are important for the general understanding of speech and gesture interactions on the neural level. Furthermore, our study ecologically validates some experimental findings about relational memory processes using a more natural setting of memory for coverbal gestures.

The recognition task was difficult for our participants, as demonstrated by the relatively low memory performances (e.g., 60% hits in the MG condition). Overall 256 Videos of eight conditions were presented during encoding. All of them (during encoding and recognition) were very similar (e.g., same actor, same context) and the participants were unaware that they would have to remember the videos after the fMRI session. These facts probably contributed to the low memory performance and the high variability across participants. Despite this low performance, the brain activation patterns suggest that encoding processes for the MG versus the FG and NG conditions are different. In addition, in the recognition phase, our participants were instructed to judge whether they had previously seen the video. Our participants could have focused only on one modality of the bimodal conditions. However, for an adequate response, it would be important to incorporate information of both modalities. Furthermore, we also see interacting effects of both modalities in differences between the evaluations of the FG and MG conditions. This most likely indicated integration of both modalities reflected in increases in ''understandability,'' ''imageability,'' and ''naturalness scores'' and at least the identification of incongruent information from both modalities in decreases of ''naturalness scores.'' Importantly, these differences are not a result of the content of speech (the same sentences were used in each condition) or gesture (no correlation between gesture comprehension in isolation and memory performance of the bimodal conditions) and therefore represent an effect of combined speech and gesture processing.

Overall, our study showed cortical and hippocampal activation in relation to the subsequent memory performance for speech and gesture information. The implicit encoding of abstract speech and metaphoric gestures showed specific contribution of a left lateral frontotemporal language network indicating a more semantic level of integration (binding). These results give novel evidence of the neural substrates involved in the formation of memory representations for speech and gesture information on an abstract level.

Acknowledgments

This research project is supported by a grant from the Interdisciplinary Center for Clinical Research ''BIOMAT'' (IZKF VV N68) and the Deutsche Forschungsgemeinschaft (DFG, IRTG 1328).

We thank Klaus Willmes for the fruitful discussion of a former draft of this manuscript and the whole IZKF service team for their allocation of necessary technical and personnel requirements.

Reprint requests should be sent to Benjamin Straube, Department of Psychiatry, RWTH Aachen University, Pauwelsstr. 30, D-52074 Aachen, Germany, or via e-mail: bstraube@ukaachen.de.

REFERENCES

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe—A simple network model. *Proceedings of the National Academy of Sciences, U.S.A., 91,* 7041–7045.

Amunts, K., Kedo, O., Kindler, M., Pieperhoff, P., Mohlberg, H., Shah, N. J., et al. (2005). Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: Intersubject variability and probability maps. *Anatomy and Embryology (Berlin), 210,* 343–352.

Baker, J. T., Sanders, A. L., Maccotta, L., & Buckner, R. L. (2001). Neural correlates of verbal memory encoding during semantic and structural processing tasks. *NeuroReport, 12,* 1251–1256.

Beattie, G., & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology, 184,* 438–462.

Beattie, G., & Shovelton, H. (2002). What properties of talk are associated with the generation of spontaneous iconic hand gestures? *British Journal of Social Psychology, 41,* 403–417.

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science, 281,* 1185–1187.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience, 13,* 400–404.

Calvo-Merino, B., Glaser, D. E., & Grezes, J. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex, 15,* 1243–1249.

Casasanto, D. J., Killgore, W. D. S., Maldjian, J. A., Glosser, G., Alsop, D. C., Cooke, A. M., et al. (2002). Neural correlates of successful and unsuccessful verbal memory encoding. *Brain and Language, 80,* 287–295.

Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system*. Cambridge: MIT Press.

Daselaar, S. M., Rombouts, S. A., Veltman, D. J., Raaijmakers, J. G., Lazeron, R. H., & Jonker, C. (2001). Parahippocampal activation during successful recognition of words: A self-paced event-related fMRI study. *Neuroimage, 13,* 1113–1120.

Davachi, L., Maril, A., & Wagner, A. D. (2001). When keeping in mind supports later bringing to mind: Neural markers of phonological rehearsal predict subsequent remembering. *Journal of Cognitive Neuroscience, 13,* 1059–1070.

Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences, U.S.A., 100,* 2157–2162.

Davachi, L., & Wagner, A. D. (2002). Hippocampal contributions to episodic encoding: Insights from relational and item-based learning. *Journal of Neuroscience, 88,* 982–990.

Donaldson, D. I., Petersen, S. E., & Buckner, R. L. (2001). Dissociating memory retrieval processes using fMRI: Evidence that priming does not support recognition memory. *Neuron, 31,* 1047–1059.

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron, 44,* 109–120.

Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain*. Oxford: Oxford University Press.

Eichenbaum, H., Otto, T., & Cohen, N. J. (1992). The hippocampus—What does it do? *Behavioral and Neural Biology, 57,* 2–36.

Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y., & Engel, S. A. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. *Nature Neuroscience, 3,* 1149–1152.

Fletcher, P. C., Stephenson, C. M., Carpenter, T. A., Donovan, T., & Bullmorel, E. T. (2003). Regional brain activations predicting subsequent memory success: An event-related fMRI study of the influence of encoding tasks. *Cortex, 39,* 1009–1026.

Fliessbach, K., Trautner, P., Quesada, C. M., Elger, C. E., & Weber, B. (2007). Cerebellar contributions to episodic memory encoding as revealed by fMRI. *Neuroimage, 35,* 1330–1337.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *Neuroimage, 7,* 30–40.

Gauger, P. W. (1952). The effect of gesture and the presence or absence of the speaker on the listening comprehension of eleventh and twelfth grade high school pupils. *Speech Monographs, 19,* 116–117.

Gold, J. J., Smith, C. N., Bayley, P. J., Shrager, Y., Brewer, J. B., Stark, C. E. L., et al. (2006). Item memory, source memory, and the medial temporal lobe: Concordant findings from fMRI and memory-impaired patients. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 9351–9356.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences, 3,* 419–429.

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.

Hanlon, R., Brown, J., & Gerstman, L. (1990). Enhancement of naming in nonfluent aphasia through gesture. *Brain and Language, 38,* 298–314.

Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: A neuromagnetic study. *Proceedings of the National Academy of Sciences, U.S.A., 95,* 15061–15065.

Henson, R. A., Rugg, M. D., Shallice, T., Josephs, O., & Dolan, R. J. (1999). Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study. *Journal of Neuroscience, 19,* 3962–3972.

Henson, R. N., Hornberger, M., & Rugg, M. D. (2005). Further dissociating the processes involved in recognition memory: An fMRI study. *Journal of Cognitive Neuroscience, 17,* 1058–1073.

Holle, H., Gunter, T. C., Rüschemeyer, S.-A., Hennenlotter, A., & Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *Neuroimage, 39,* 2010–2024.

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.

Kirchhoff, B. A., & Buckner, R. L. (2006). Functional–anatomic correlates of individual differences in memory. *Neuron, 51,* 263–274.

Kirchhoff, B. A., Wagner, A. D., Maril, A., & Stern, C. E. (2000). Prefrontal–temporal circuitry for episodic encoding and subsequent memory. *Journal of Neurosciences, 20,* 6173–6180.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 262,* 23–81.

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

McNeill, D. (1999). Action, thought and language. In P. Llyod & C. Fernyhough (Eds.), *Lev Vygotsky: Critical assessments: Thought and language* (Vol. II, pp. 23–30). Florence, KY: Taylor & Francis/Routledge.

Morcom, A. M., Good, C. D., Frackowiak, R. S., & Rugg, M. D. (2003). Age effects on the neural correlates of successful memory encoding. *Brain, 126,* 213–229.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary learning-systems approach. *Psychological Review, 110,* 611–646.

Nyberg, L., Mcintosh, A. R., Houle, S., Nilsson, L. G., & Tulving, E. (1996). Activation of medial temporal structures during episodic memory retrieval. *Nature, 380,* 715–717.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97–113.
Otten, L. J., Henson, R. N., & Rugg, M. D. (2001). Depth of processing effects on neural correlates of memory encoding: Relationship between findings from across- and within-task comparisons. *Brain, 124,* 399–412.
Otten, L. J., & Rugg, M. D. (2001). Task-dependency of the neural correlates of episodic encoding as measured by fMRI. *Cerebral Cortex, 11,* 1150–1160.
Ozyurek, 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.
Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences, 6,* 93–102.
Prince, S. E., Daselaar, S. M., & Cabeza, R. (2005). Neural correlates of relational memory: Successful encoding and retrieval of semantic and perceptual associations. *Journal of Neurosciences, 25,* 1203–1210.
Prince, S. E., Tsukiura, T., & Cabeza, R. (2007). Distinguishing the neural correlates of episodic memory encoding and semantic memory retrieval. *Psychological Science, 18,* 144–151.
Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience, 4,* 193–202.
Rolls, E. T. (2000). Memory systems in the brain. *Annual Review of Psychology, 51,* 599–630.
Rose, M., & Douglas, J. (2001). The differential facilitatory effects of gesture and visualisation processes on object naming in aphasia. *Aphasiology, 15,* 977–990.
Shastri, L. (2002). Episodic memory and cortico-hippocampal interactions. *Trends in Cognitive Neurosciences, 6,* 162–168.
Slotnick, S. D., Moo, L. R., Segal, J. B., & Hart, J., Jr. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Brain Research, Cognitive Brain Research, 17,* 75–82.
Strange, B. A., Otten, L. J., Josephs, O., Rugg, M. D., & Dolan, R. J. (2002). Dissociable human perirhinal, hippocampal, and parahippocampal roles during verbal encoding. *Journal of Neuroscience, 22,* 523–528.
Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
Uncapher, M. R., & Rugg, M. D. (2005). Effects of divided attention on fMRI correlates of memory encoding. *Journal of Cognitive Neuroscience, 17,* 1923–1935.
Valenzeno, L., Alibali, M. W., & Klatzkya, R. (2003). Teacher's gestures facilitate students learning: A lesson in symmetry. *Contemporary Educational Psychology, 28,* 187–204.
Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., et al. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science, 281,* 1188–1191.
Wallenstein, G. V., Eichenbaum, H., & Hasselmo, M. E. (1998). The hippocampus as an associator of discontiguous events. *Trends in Neurosciences, 21,* 317–323.
Weis, S., Klaver, P., Reul, J., Elger, C. E., & Fernandez, G. (2004). Temporal and cerebellar brain regions that support both declarative memory formation and retrieval. *Cerebral Cortex, 14,* 256–267.
Wheeler, M. E., & Buckner, R. L. (2004). Functional–anatomic correlates of remembering and knowing. *Neuroimage, 21,* 1337–1349.
Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language, 101,* 278–289.
Willems, R. M., Ozyurek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex, 17,* 2322–2333.
Witter, M. P., & Amaral, D. G. (1991). Entorhinal cortex of the monkey: V. Projections to the dentate gyrus, hippocampus, and subicular complex. *Journal of Comparative Neurology, 307,* 437–459.
Woodruff, C. C., Johnson, J. D., Uncapher, M. R., & Rugg, M. D. (2005). Content-specificity of the neural correlates of recollection. *Neuropsychologia, 43,* 1022–1032.