Behavioral/Systems/Cognitive

What You See Is Not (Always) What You Hear: Induced Gamma Band Responses Reflect Cross-Modal Interactions in Familiar Object Recognition

Shlomit Yuval-Greenberg[1] and Leon Y. Deouell[1,2]

[1]Department of Psychology, and [2]Interdisciplinary Center for Neural Computation, The Hebrew University of Jerusalem, Jerusalem 91905, Israel

Gamma-band responses (GBRs) are hypothesized to reflect neuronal synchronous activity related to activation of object representations. However, it is not known whether synchrony in the gamma range is also related to multisensory object processing. We investigated the effect of semantic congruity between auditory and visual information on the human GBR. The paradigm consisted of a simultaneous presentation of pictures and vocalizations of animals, which were either congruent or incongruent. EEG was measured in 17 students while they attended either the auditory or the visual stimulus and performed a recognition task. Behavioral results showed a congruity effect, indicating that information from the unattended modality affected behavior. Irrelevant visual information affected auditory recognition more than irrelevant auditory information affected visual recognition, suggesting a bias toward reliance on visual information in object recognition. Whereas the evoked (phase-locked) GBR was unaffected by congruity, the induced (non-phase-locked) GBR was increased for congruent compared with incongruent stimuli. This effect was independent of the attended modality. The results show that integration of information across modalities, based on semantic congruity, is associated with enhanced synchronized oscillations at the gamma band. This suggests that gamma-band oscillations are related not only to low-level unimodal integration but also to the formation of object representations at conceptual multisensory levels.

Key words: auditory; event-related potentials; EEG; gamma; human; multisensory; object recognition; oscillations; visual

Introduction

How does the brain integrate the activity of diverse populations of neurons coding different aspects of the same entity into a unified percept? One plausible account suggests that disparate neuronal populations form synchronized oscillating networks. Specifically, gamma-band oscillations (~30 Hz and up) were suggested to reflect this mechanism (Başar-Eroğlu et al., 1996). Thus, it was found that cat primary visual neurons coding different parts of a line fire synchronously at the gamma range (Eckhorn et al., 1988; Gray and Singer, 1989), and comparable findings were found in human EEG, suggesting an even larger scale synchrony (Lutzenberger et al., 1995; Müller et al., 1996).

The human EEG gamma-band response (GBR) usually appears as two components: an early evoked (phase-locked) response ~100 ms after stimulus onset (eGBR) and a later induced (non-phase-locked) response ~200–300 ms after stimulus onset (iGBR) (Tallon-Baudry et al., 1996). In a seminal series of studies, Tallon-Baudry et al. showed that human iGBR is augmented in response to images that create the perceptual illusion of identifiable objects relative to images that do not (Tallon et al., 1995; Tallon-Baudry et al., 1996, 1997, 1998). This effect is correlated with the subjects' ability to consciously detect the object and with its familiarity (Lutzenberger et al., 1994; Tallon-Baudry et al., 1996, 1997; Gruber et al., 2002; Fiebach et al., 2005; Gruber and Müller, 2005; Busch et al., 2006; Lenz et al., 2007). Whereas the iGBR response may be too late to reflect the low-level binding of object parts (Murray et al., 2002, 2004), these studies suggested that increased iGBR is associated with the activation of a mental object representation (Tallon-Baudry and Bertrand, 1999). Congruently, scalp EEG and intracranial recordings in humans support distributed, long-range synchronization underlying the iGBR in response to visual objects (Gruber and Müller, 2005; Lachaux et al., 2005).

Most of this research has so far focused on unimodal stimuli, mainly in the visual modality. However, in reality, events have multisensory dimensions. A car has a characteristic sound, as does a dog. Are the electrophysiological manifestations of integration across modalities similar to those seen within a modality? More particularly, is iGBR involved in the recognition of a multisensory integrated object and does coherence at the semantic level affect neural synchrony?

Previously, multisensory investigations of the GBR were limited to integration of elementary meaningless stimuli [tones and gratings (Senkowski et al., 2005, 2007)] or to lip-reading phenomena (Kaiser et al., 2006) known to be low-level processes (McGurk and MacDonald, 1976; Sams et al., 1991). Here, we investigated the effect of two semantically meaningful sources of information on the GBR by simultaneously presenting pictures and vocalizations of animals, which were either congruent (the
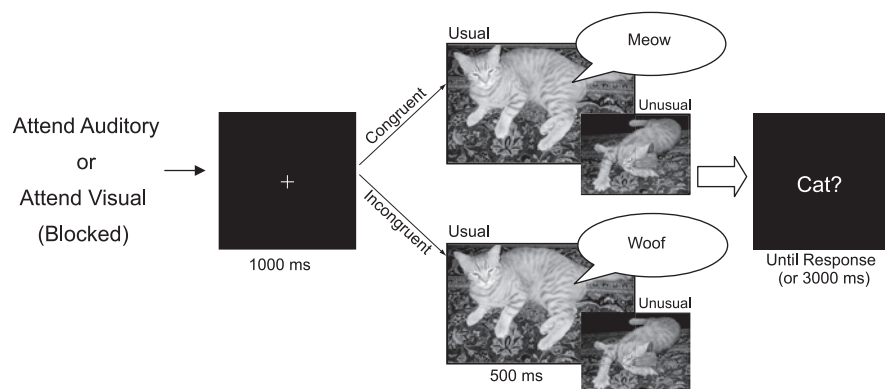
Figure 1. Trial procedure. The task was predetermined by a question presented at the beginning of each block: "Which animal did you see?" or "Which animal did you hear?" The fixation cross was presented for 1 s, followed immediately by one of the stimulus combinations for 500 ms, which was followed by a forced-choice yes/no question (e.g., "dog?") until a response was given or until 3 s elapsed.

same animal) or incongruent (different animals). Subjects were asked to identify either the picture or the vocalization and to ignore the other modality. Cross-modal effects were expected to facilitate responses and to increase the iGBR when the visual and the auditory objects were congruent, relative to when they were incongruent.

Materials and Methods

Subjects. Twenty-four students of The Hebrew University of Jerusalem participated in the experiment in exchange for a fee or for course credits. All reported normal hearing and normal or corrected-to-normal vision with no history of neurological problems. Seven subjects were excluded from analysis: five because of excessive EEG artifacts (more than two-thirds of the trials had to be rejected) and two because they failed to perform the task. Thus, behavior and event-related potentials (ERPs) were analyzed from 17 subjects (eight females, nine males; mean age, 23.7 years; 16 right handed). The study was conducted in accordance with the ethical regulations of The Hebrew University of Jerusalem. All subjects signed an informed consent.

Stimuli. The auditory stimuli included 48 vocalizations of six familiar animals (cat, dog, rooster, bird, sheep, and cow; eight different exemplars for each animal). Sounds were presented from a loudspeaker placed directly below the monitor. Pilot studies verified that these vocalizations could be recognized with very high accuracy. Visual stimuli consisted of color photographs of the same animals (eight for each of the six animals). Half of the pictures were of animals in prototypical views and in their natural setting ("usual condition"). These pictures were very easy to recognize. The remaining pictures were harder to recognize because they were taken from unusual angles or were of animals in unusual postures (e.g., a dog standing on hindlegs photographed from below; "unusual condition"). The inclusion of the usual and unusual pictures was based on preliminary behavioral studies (S. Yuval-Greenberg, Y. Avivi, and L. Y. Deouell, unpublished observation), which showed that irrelevant prototypical pictures affected recognition of sounds more than vice versa. We hypothesized that this might be because the prototypical visual stimuli were easier to recognize and thus included here a set of "unusual" pictures. Two pilot studies were conducted to ensure that the unusual pictures were indeed more difficult to recognize than the "usual" pictures. In these experiments, both usual and unusual pictures were presented along with a neutral auditory stimulus (series of clicks). The order of the pictures was random, and each picture was presented four times during the course of the experiment. Subjects were asked either to choose which of six animals appeared in the picture by pressing one of six buttons (forced-choice task, five subjects) or to name the animal in the picture (naming task, 10 subjects). Naming reaction times were measured with a voice-key apparatus.

Procedure. Subjects were seated in a sound-attenuated chamber. The experiment included four main blocks (attend–auditory, usual; attend–auditory, unusual; attend–visual, usual; and attend–visual, unusual) of 192 trials each. The order of the blocks was counterbalanced across subjects. The trials in all blocks consisted of the concurrent presentation of a visual and an auditory stimulus. In the usual blocks, the visual stimuli consisted of the easier-to-recognize prototypical pictures, and, in the unusual blocks, the more difficult pictures were presented. In addition, two short training blocks, each including 12 trials and feedback on performance, were presented before the appropriate (visual/auditory) main block. The pictures and vocalizations used in the training block were not presented in the main blocks.

The task was a modification of the Name Verification Paradigm (Thompson-Schill et al., 1997). Each block started with an instruction slide specifying the current task. In the two attend–auditory blocks, subjects were instructed to recognize and respond based on the auditory stimulus alone. In the corresponding attend–visual blocks, the subjects were instructed to recognize and respond based on the visual stimulus alone. In all blocks, half of the trials were "congruent" (picture and voice of the same animal) and half were "incongruent" (picture and voice belonging to different animals). The order of the trials was random. Each picture and each vocalization was presented four times in congruent trials and four times in incongruent trials. Because the number of possible incongruent combinations is very large, a different list of incongruent trials was randomly selected for each subject.

Trials started with a fixation cross presented for 1 s (Fig. 1). The audiovisual stimulus was then presented, lasting for 500 ms, followed immediately by a forced-choice yes/no question (e.g., "dog?"), which remained on the screen until either a response was given or 3 s had elapsed. In the incongruent trials, the question always corresponded to either the visual or the auditory stimulus. Subjects responded by pressing one of two buttons to indicate their response. The correct response was affirmative in half of the trials and negative in the other half, independent of whether the trial was congruent or incongruent. The reason for deferring the yes/no question (i.e., the word to be verified) until after the presentation of the audiovisual stimulus was to delay the response selection process and thus disentangle the brain responses related to perceptual conflicts from those related to response conflicts. Note that, up to the presentation of the question, some 500 ms after the stimulus was presented, response selection processes could not commence.

EEG recording. Recording was done with an Active 2 acquisition system (BioSemi, Amsterdam, The Netherlands) at a sampling rate of 1024 Hz with on-line low-pass filter of 268 Hz to prevent aliasing of high frequencies. Recording included 64 sintered Ag/AgCl electrodes placed on a cap according to the extended 10–20 system and seven additional electrodes [two on the mastoids, two horizontal electro-oculography (EOG) channels, two vertical EOG channels, and a nose channel].

Behavior analysis. For each subject, trials with reaction times of >2 SDs from the mean of each condition were discarded. Only reaction times of correct responses were analyzed. Differences between conditions were evaluated using a three-way repeated-measures ANOVA with factors attended modality (visual, auditory), congruity (congruent, incongruent), and picture typicality (usual, unusual).

EEG analysis. The EEG was analyzed using Brain Vision Analyzer software (version 1.05; Brain Products, Gilching, Germany). Data were referenced to the nose channel and digitally high-pass filtered at 0.5 Hz (24 dB/octave), using a Butterworth zero-shift filter. Activity attributable to blinks was removed from the data before segmentation, using the independent component analysis method (Jung et al., 2000). In addition, epochs with gross eye movements and other artifacts were rejected using an amplitude criterion of ±100 μV. The total number of trials analyzed per subject per condition was 46–96 (of 96), with a mean of 86.

The time-frequency analysis was done by convolving data with a complex Morlet's wavelet with a Gaussian shape in both the time domain (SD σ_t) and the frequency domain (SD σ_f) around its central frequency f_0: $W(t,f_0) = A \exp(-t^2/2\sigma_t^2)\exp(2i\pi f_0 t)$ using a constant ratio of $f/\sigma_f = 8$, where $\sigma_f = 1/(2\pi\sigma_t)$ and normalization factor of A (Tallon-Baudry et al., 1996; Herrmann et al., 1999). This procedure was applied to frequencies ranging from 20 to 100 Hz in steps of 0.675 Hz. The wavelet calculation was performed on epochs starting from 300 ms before stimulus onset until 700 ms after stimulus onset and was baseline corrected, using the baseline range of -200 to -100 ms before stimulus onset. (Because by its nature the convolution with wavelets introduces time smearing, using a baseline that ends at 0, as done for ERPs, would "contaminate" the baseline with activity, which in fact happened after stimulus onset. This procedure is the rule in time-frequency analyses.) Evoked (phase-locked) activity was calculated for each subject by applying the wavelet analysis to the average of the epochs. This procedure is insensitive to induced oscillations because of their intertrial phase jitter. A total of both induced and evoked activity was calculated by performing the wavelet analysis on single trials and only then averaging the resulting single-trial frequency spectra. An unbiased frequency × time × channels region of interest (ROI) was defined for the evoked and induced components separately, based on the region of maximal activity found in the average of all conditions combined. [Several studies performed gamma analysis on the maximum frequencies of activity chosen individually for each subject. Because in the current study, the activation of the induced GBR was wide in frequency, an ROI analysis was used (for a similar analysis, see Gruber and Müller, 2002).] The existence of the early and the late components was confirmed by comparing the GBR across all experimental conditions with the baseline. The average power within this ROI for each condition was then used as the dependent variable in a three-way ANOVA with factors picture typicality (usual, unusual), attended modality (visual, auditory), and congruity (congruent, incongruent).

ERPs resulted from averaging segments that lasted from 200 ms before stimulus onset until 700 ms after stimulus onset, digitally filtered with a bandpass of 1–17 Hz (24 dB/octave) using a Butterworth zero-shift filter, and referenced to the mean of the prestimulus period. The amplitude of the visual N1 served as a predetermined dependent variable based on a previous result (Molholm et al., 2004). In addition, two *post hoc* windows were measured based on prominent peaks in the resulting waveforms (see Results). Finally, to explore the data point-by-point for effects at unexpected latencies, while controlling for inflation of type I error, we downsampled the data to 256 Hz, limited the epoch to 0–500 ms (i.e., before the question appears), and used a nonparametric permutation analysis to determine the t threshold yielding an experiment-wise error of <0.05 (Blair and Karniski, 1993). The details of this method are explained in the supplemental data (available at www.jneurosci.org as supplemental material).

Results
Behavior
Pilot experiments
In the unimodal pilot experiments, conducted to compare the difficulty of recognizing usual and unusual pictures, the subjects responded faster to usual than to unusual pictures in the forced-choice task ($t_{(4)} = 2.8$; $p < 0.05$) and were marginally faster for usual pictures in the naming task ($t_{(9)} = 1.71$; $p = 0.06$). Notably, naming the usual pictures in their first presentation was significantly ($t_{(9)} = 6.02$; $p < 0.0001$) faster than naming the unusual pictures on first presentation. The results of both experiments confirmed that the unusual pictures were indeed harder to recognize than the usual pictures.

Main experiment
Subjects were faster in the attend–visual task than in the attend–auditory task. They were also faster and more accurate in responding to congruent than to incongruent trials. A three-way ANOVA was conducted, with factors typicality (usual, unusual), attended modality (visual, auditory), and congruity (congruent,

incongruent). This analysis revealed that subjects were overall faster to respond in the attend–visual task (732 ms) than in the attend–auditory task (799 ms; main effect of modality; $F_{(1,16)} = 14.2$; $p < 0.005$). Congruity also affected performance significantly (congruent, 739 ms; incongruent, 792 ms; main effect of congruity $F_{(1,16)} = 62.59$; $p < 0.00001$). However, the effect of congruity was weaker in the attend–visual task (incongruent–congruent, 39 ms) than in the attend–auditory task (67 ms), as indicated by the significant modality × congruity interaction ($F_{(1,16)} = 6$; $p < 0.05$). There was no main effect of typicality ($F_{(1,16)} < 1$) and no significant interaction between typicality and any of the other factors (all $F_{(1,16)} < 1$).

Accuracy rates did not demonstrate a main effect of modality ($F_{(1,16)} < 1$). Subjects were more accurate in the congruent condition than in the incongruent condition (main effect of congruity, $F_{(1,16)} = 27.7$; $p < 0.0001$). This effect was stronger in the attend–auditory (7%) than in the attend–visual task (2%; modality × congruity interaction, $F_{(1,16)} = 7.76$; $p < 0.05$), precluding a speed–accuracy tradeoff. There was no main effect of typicality ($F_{(1,16)} = 3.47$; $p = 0.08$) and no significant interactions between typicality and any of the other factors (typicality × modality, $F_{(1,16)} < 1$; typicality × congruity, $F_{(1,16)} = 2.37$; $p = 0.14$; typicality × modality × congruity, $F_{(1,16)} < 1$).

Gamma-band activity
Evoked GBR
Visual inspection of the evoked time-frequency plots of the gamma range averaged across all conditions revealed an evoked gamma component peaking at ~90 ms. Based on the average activity across conditions, a window of 70–110 ms and 20–50 Hz was chosen as an unbiased ROI for the analysis. This activity was sampled in two clusters of electrodes showing the maximal activity across the scalp (posterior: PO7, PO8, O1, O2, Oz, P9, P10, and Iz; central: CPz, CP1, CP2, Cz, C1, C2, FCz, FC1, and FC2) (Fig. 2). The average power within these time × frequency × channel ROIs served as the dependent variable for the subsequent analyses of eGBR.

A robust eGBR was found in each condition (Fig. 2). The increase in eGBR in this range was found significant compared with 0 in all conditions in both ROIs ($p < 0.05$ for all conditions). A three-way ANOVA found, however, no significant effects of typicality (posterior ROI, $F_{(1,16)} = 4.38$, $p = 0.053$; central ROI, $F_{(1,16)} = 1.34$, $p = 0.26$), attended modality ($F_{(1,16)} < 1$ for both ROIs), or congruity ($F_{(1,16)} < 1$ for both ROIs), nor any significant interactions between these factors ($F_{(1,16)} < 1$ for all). Similar results were obtained when analysis was done on the single most active channel (PO7) instead of the clusters.
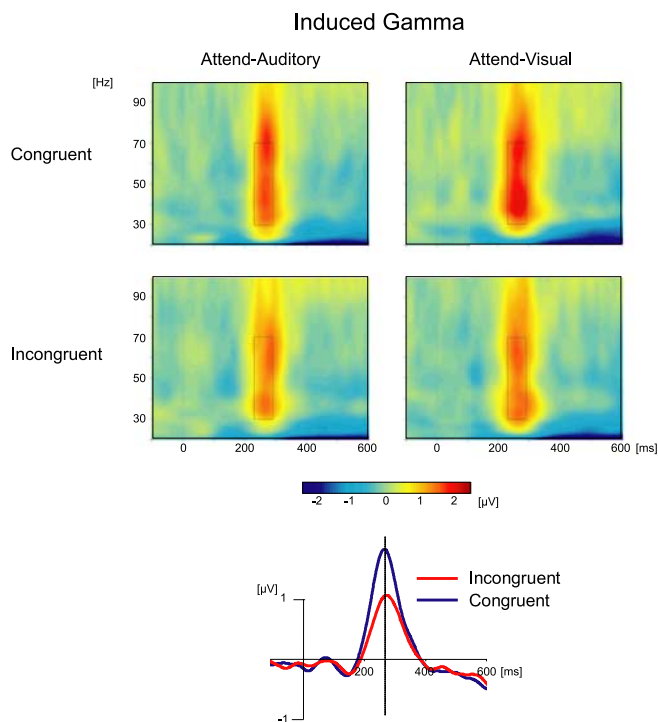
Induced GBR
Visual inspection of the time-frequency plot of the iGBR averaged across all conditions combined revealed a prominent increase peaking at 260 ms. This increase had a wide scalp distribution with a peak at the posterior–midline region. Based on this unbiased average, the analysis ROI was chosen as a window of 230–290 ms and 30–70 Hz at a cluster of eight occipitoparietal channels (Pz, POz, P1, P2, P3, P4, PO3, and PO4) showing the maximal activity. The average power within this ROI served as the dependent variable for the subsequent analyses of iGBR.

An iGBR peak at the ROI range was seen in each condition (Fig. 3). This increase was found significant compared with 0 ($p < 0.01$) in all conditions. Three-way ANOVA with factors typicality (usual, unusual), attended modality (auditory, visual), and congruity (congruent, incongruent) revealed a significant

Evoked Gamma
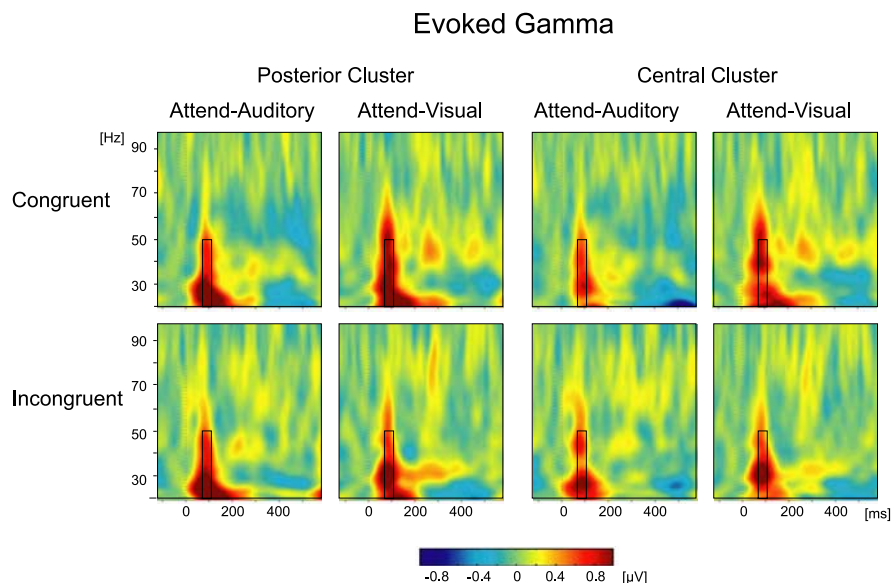
Figure 2. Evoked GBR. Time-frequency plots produced by performing wavelet transformation on the average response. Plots are shown for both tasks and both congruity conditions and for the two clusters of electrodes (posterior: P07, P08, 01, 02 , 0z, P9, P10, and Iz; central: CPz, CP1, CP2, Cz, C1, C2, FCz, FC1, and FC2), collapsed across the usual and unusual pictures. The black rectangle denotes the time × frequency analysis ROI.

Induced Gamma

Figure 3. Induced GBR. Top, Time-frequency plots produced by averaging wavelet transformation of single trials. Plots are shown for both tasks and both congruity conditions, collapsed across the usual and unusual picture conditions, which had no effect, and across a cluster of electrodes (Pz, P0z, P1, P2, P3, P4, PO3, and PO4). The black rectangle denotes the time × frequency analysis ROI. Note the larger iGBR in congruent than incongruent conditions. Bottom, Induced GBR in congruent and incongruent trials represented by a single wavelet, centered at 40 Hz. The data were collapsed across picture typicality and task, which had no significant effects.

main effect of congruity ($F_{(1,16)} = 14.37$; $p < 0.005$). The effect was attributable to an enhanced iGBR in the congruent condition compared with the incongruent condition (Fig. 3). The interactions of congruity and attended modality ($F_{(1,16)} = 1.08$; $p = 0.31$) and of congruity and typicality ($F_{(1,16)} = 1.37$; $p = 0.26$) were not

significant. All other main effects and interactions were not significant as well (typicality main effect, $F_{(1,16)} = 2.28$, $p = 0.15$; attended modality main effect, $F_{(1,16)} = 2.07$, $p = 0.17$; modality × typicality, $F_{(1,16)} = 1.7$, $p = 0.21$; congruity × modality × typicality, $F_{(1,16)} < 1$). Similar results were obtained when analysis was done on the single most active channel (POz) instead of a cluster. The scalp distribution of the iGBR was similar across conditions (Fig. 4).

Event-related potentials
Inspection of the waveforms elicited by congruent and incongruent trials did not reveal conspicuous differences in either task (Fig. 5) (supplemental Figs. 1, 2, available at www.jneurosci.org as supplemental material). Based on previous results (Molholm et al., 2004), we looked for effects of congruity, as well as attended modality and typicality, at the latency and distribution of the occipitotemporal N1 (peak latency, 175 ms). [We consider this a visual N1, because a component with similar latency and distribution was elicited only in the visual condition of a preliminary study in which 14 subjects performed the same task with unimodal auditory or visual stimuli (Yuval-Greenberg, Avivi, and Deouell, unpublished observation).] A four-way ANOVA (modality × congruity × typicality × hemisphere) on the mean amplitude of 20 ms around the N1 peak at electrodes P7/8, P9/10, and PO7/8 revealed no effect of congruity or interaction of congruity with any other factor (all $F_{(1,16)} < 1$ except congruity × modality, in which $F_{(1,16)} = 2.37$, $p = 0.14$). The N1 tended to be stronger on the left ($F_{(1,16)} = 1.71$; $p = 0.2$), and this tendency was stronger in the attend–auditory than the attend–visual condition (hemisphere × modality interaction, $F_{(1,16)} = 8.54$; $p < 0.05$). For completeness, we also conducted exploratory analyses (see Materials and Methods) to look for congruity effects in other latencies or loci but found no significant differences. Additional details on this analysis are provided in the supplemental data (available at www.jneurosci.org as supplemental material).

Discussion
We investigated the effects of semantically congruent and incongruent multisensory inputs during object recognition on human gamma-band responses. We hypothesized that, if iGBR is related to the formation of coherent object representations, as has been suggested for the unimodal case, then it could also be related to larger-scale, multisensory object representations. Compatibly, we find that cross-modal congruity is associated with improved object recognition and with an increase in the induced GBR.

Behavioral indices of cross-modal congruity
Previous results examining elementary spatial, temporal, and linguistic interactions showed that the cross-modal interaction is involuntary (Colin et al., 2002; Bertelson and de Gelder, 2004; Stekelenburg et al., 2004). The current results suggest that this is true also for more complex object recognition.

Although the behavioral "congruity effect" appeared both when the task was visual and when it was auditory, it was larger

for the latter. Thus, in the context of object recognition, the visual modality influences auditory processing more than vice versa, suggesting that the visual system may be the "appropriate" modality for object recognition. By the "modality appropriateness" model of cross-modal integration (Welch and Warren, 1980), the dominant modality in case of conflict is the modality better equipped for processing a specific stimulus dimension (e.g., vision for spatial processing). Recent findings, however, show that the influence from the nominally "inferior" modality might be more evident when the information in the usually dominant modality is deficient (Ernst et al., 2000; Ernst and Banks, 2002; Hairston et al., 2003a,b; Heron et al., 2004; Witten and Knudsen, 2005). Notwithstanding, the typicality of pictures in our experiment did not affect the asymmetry of the congruity effect. Thus, it seems that, in face of conflict in object recognition, our cognitive system is strongly biased toward visual information. However, our manipulation of visual difficulty may not have been robust enough to reveal information effects. As the pilot tests showed, learning could have reduced the difference between the two sets of images along the experiment.

Neurophysiological indices of cross-modal congruity

The multisensory tasks elicited both an early eGBR ~90 ms and a later iGBR ~260 ms after stimulus onset. These latencies are consistent with previous results using unimodal presentations (Tallon-Baudry et al., 1996; Gruber and Müller, 2005). The crux of the present findings is in the effect of cross-modal congruity on this response. Whereas there was no significant effect on the eGBR, the later iGBR was enhanced when multisensory stimuli were semantically congruent compared with incongruent. This multisensory iGBR effect is, to our knowledge, the first evidence for the generality of the iGBR as an indicator for coherent object representation across modalities.

Enhanced iGBR might reflect neural synchronous activity at several levels of object representation (Başar-Eroğlu et al., 1996). Previously, enhanced iGBR was seen for elementary stimuli, which conformed to basic Gestalt laws, relative to those that did not (Eckhorn et al., 1988; Lutzenberger et al., 1995), for parts that created the illusion of a recognizable object more than for parts that did not (Tallon-Baudry et al., 1997), and for familiar objects with long-term memory representations compared with novel objects (Herrmann et al., 2004; Gruber and Müller, 2005; Lenz et al., 2007). In addition, it was argued that iGBR enhancements could reflect "target effects," that is, resemblance between stimuli and a target held in working memory (Tallon-Baudry et al., 1998).

At what level could the present results be explained? Target effects are unlikely because all stimuli were equally targeted in our paradigm. Low-level unimodal "Gestalt effects" can be ruled out as well, because all unimodal stimuli constituting the multisensory presentations formed good Gestalts. Moreover, the latency of the iGBR, peaking ~260 ms, seems too late to reflect low-level binding of object parts, and, in fact, Murray et al. (2002) have
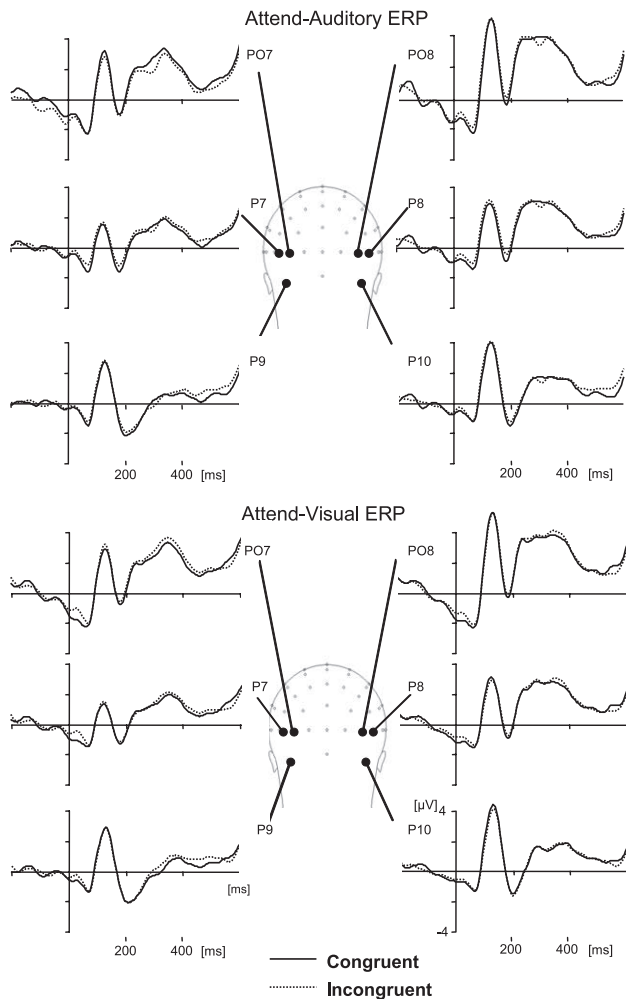
Induced Gamma Distributions
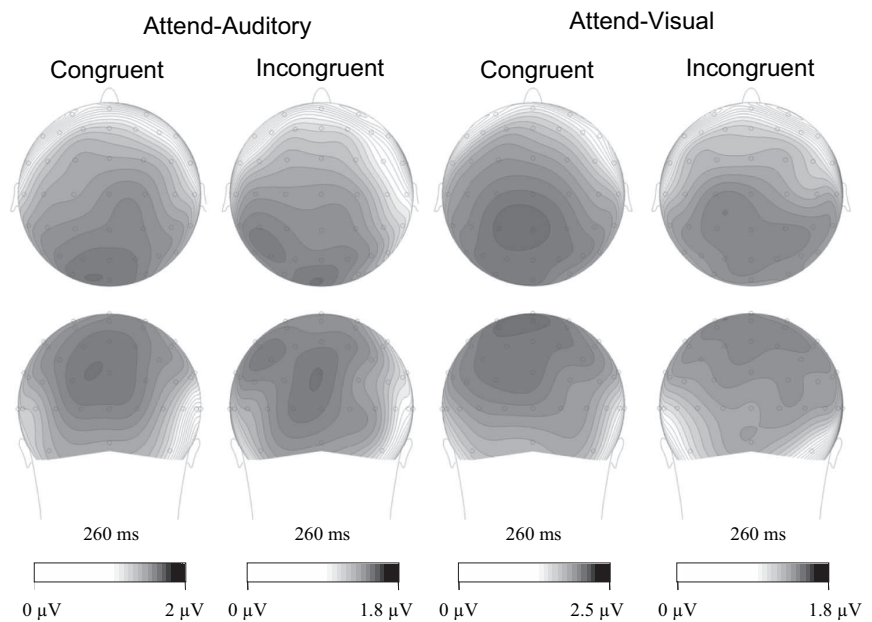
Figure 4. Topographical distributions of the induced GBR in both task and both congruity conditions. Note that the scales of the scalp distributions were adjusted for each condition separately to highlight the similarity of the spatial distribution of the iGBR independent of its intensity.

shown that the presence of an illusory Kanizsa shape modulates visual ERPs as early as 90 ms. The separate activation of unimodal object representations or unimodal long-term memory traces cannot explain the present effect either, because all auditory and visual stimuli were of familiar animals with representations in memory. Rather, the iGBR congruity effect had to stem from the interaction between the multiple sources of information presented in the different modalities.

We propose that this multisensory effect could be based at several levels. First, the enhanced iGBR might reflect the formation of a multisensory object representation. According to this explanation, feedback from semantic nodes would synchronize neurons in object-related visual and auditory sensory cortices, or in multisensory zones, to form a coherent neuronal assembly. Second, the effect might originate from a match between the presented stimuli and stored memory traces (Gruber and Müller, 2005; Lenz et al., 2007). This is attributable to the fact that multisensory stimuli could have been represented as coherent entities in long-term memory if they were congruent but not otherwise (assuming that barking roosters are not commonly encountered). This would suggest the existence of true multisensory memory traces. Consistent with this notion, Lehmann and Murray (2005) showed that memory for object drawings was improved if the initial encounter with the drawing was accompanied by a congruent sound (relative to no sound at all) but not impaired if the initial encounter was accompanied by an incongruent sound. They suggested that this was attributable to common activation of singular multisensory representations during encoding in the congruent case and parallel activation of distinct representations (and thus no interference) in the incongruent case. Last, the enhanced iGBR could reflect the activation of neuronal ensembles forming higher-order conceptual or semantic representations. Arguably, the activation at a conceptual level would be more coherent when both stimuli activate the same rather than disparate semantic entities. Notably, these explana-

Figure 5. Grand average ERPs of occipitotemporal channels showing the N1 component, in the attend–visual and the attend–auditory tasks. There were no significant differences between the congruent (solid lines) and incongruent (dotted lines) conditions.

Sälejärvi et al., 2002; Talsma and Woldorff, 2005) (for review, see Schroeder and Foxe, 2005). In the single study that used electrophysiology to examine cross-modal semantic congruity in object recognition (line drawings and vocalizations of animals), congruent input enhanced the visual ERP components N1 and "selection negativity" (SN) (at ~150 and 280 ms, respectively) relative to incongruent input (Molholm et al., 2004). Significant differences between the paradigms may explain why we did not replicate these ERP effects. [We did not directly look for SN effects because there were no targets versus nontargets to compare in our study. However, the exploratory bootstrap analysis did not reveal any effects in the SN expected time range.] The subjects of Molholm et al. were required to attend to both modalities and to detect a specific target animal, which could appear in the visual, the auditory, or both modalities. In this situation, subjects could tune to the conspicuous physical features of the target in both modalities. This partial information might have been enough to allow for early interactions. In contrast, our subjects could not anticipate a specific target, and thus the interaction might have been delayed to a later latency when both unimodal stimuli have already been fully recognized. Moreover, the N1 and SN effects of Molholm et al. were restricted to target stimuli (i.e., no congruity effects were found for pairs of nontargets). Thus, some of these effects could be explained not only by semantic congruity but also by the fact that, unlike incongruent targets, congruent target stimuli constituted a redundant (bimodal) target. In contrast, in our study, there was no predefined target, and only one modality was relevant at a time. Nevertheless, because this report concentrates on the GBR effects and because the present ERP analysis elicited a null result, we defer the resolution of these issues to future studies.

To conclude, our results show that integration of information across modalities, based on semantic congruity, is associated with enhanced synchronized oscillations at the gamma band. This suggests that gamma-band oscillations are related not only to low-level feature integration, as seen in animal studies (Gray and Singer, 1989), but also to the formation of object representations at conceptual multisensory levels.

tions are not mutually exclusive (Tallon-Baudry, 2003). In fact, scalp iGBR might be stronger the more levels of coherent representation are involved. Regardless of the level involved, the present results clearly show that the semantic congruity of information, presented simultaneously through different modalities, is associated with increased neural oscillations at the gamma range.

In contrast to the behavioral results, the iGBR congruity effect was independent of the attended modality. This discrepancy suggests that the effects may originate from partially different mechanisms. Task performance relies heavily on mechanisms of selective attention to resolve the competition between two different objects. The results suggest that this competition is strongly biased toward the visual modality regardless of the distinctiveness of the visual exemplar. Conversely, the iGBR seems to reflect the automatic activation of a coherent multisensory or semantic representation independent of selective attention or response selection.

Although the earliest congruity effect in this study was later than 200 ms (the iGBR), this should not be taken as the earliest possible point for cross-modal interactions. Regardless of semantic congruity, bimodal stimulation per se results in early interactive effects that cannot be explained by simple summation of activity related to processing of each modality separately (Giard and Peronnet, 1999; Fort et al., 2002; Molholm et al., 2002; Teder-

References

Başar-Eroğlu C, Strüber D, Schurmann M, Stadler M, Başar E (1996) Gamma-band responses in the brain: a short review of psychophysiological correlates and functional significance. Int J Psychophysiol 24:101–112.

Bertelson P, de Gelder B (2004) The psychology of multimodal perception. In: Crossmodal space and crossmodal attention (Spence C, Driver J, eds), pp 141–177. Oxford: Oxford UP.

Blair RC, Karniski W (1993) An alternative method for significance testing of wave-form difference potentials. Psychophysiology 30:518–524.

Busch NA, Herrmann CS, Müller MM, Lenz D, Gruber T (2007) A cross-laboratory study of event-related gamma activity in a standard object recognition paradigm. NeuroImage 33:1169–1177.

Colin C, Radeau M, Soquet A, Dachy B, Deltenre P (2002) Electrophysiology of spatial scene analysis: the mismatch negativity (MMN) is sensitive to the ventriloquism illusion. Clin Neurophysiol 113:507–518.

Eckhorn R, Bauer R, Jordan W, Brosch M, Kruse W, Munk M, Reitboeck HJ (1988) Coherent oscillations: a mechanism of feature linking in the visual cortex? Multiple electrode and correlation analyses in the cat. Biol Cybern 60:121–130.

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433.

Ernst MO, Banks MS, Bulthoff HH (2000) Touch can change visual slant perception. Nat Neurosci 3:69–73.

Fiebach CJ, Gruber T, Supp GG (2005) Neuronal mechanisms of repetition priming in occipitotemporal cortex: spatiotemporal evidence from functional magnetic resonance imaging and electroencephalography. J Neurosci 25:3414–3422.

Fort A, Delpuech C, Pernier J, Giard MH (2002) Early auditory-visual interactions in human cortex during nonredundant target identification. Brain Res Cogn Brain Res 14:20–30.

Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11:473–490.

Gray CM, Singer W (1989) Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. Proc Natl Acad Sci USA 86:1698–1702.

Gruber T, Müller MM (2002) Effects of picture repetition on induced gamma band responses, evoked potentials, and phase synchrony in the human EEG. Brain Res Cogn Brain Res 13:377–392.

Gruber T, Müller MM (2005) Oscillatory brain activity dissociates between associative stimulus content in a repetition priming task in the human EEG. Cereb Cortex 15:109–116.

Gruber T, Müller MM, Keil A (2002) Modulation of induced gamma band responses in a perceptual learning task in the human EEG. J Cogn Neurosci 14:732–744.

Hairston WD, Laurienti PJ, Mishra G, Burdette JH, Wallace MT (2003a) Multisensory enhancement of localization under conditions of induced myopia. Exp Brain Res 152:404–408.

Hairston WD, Wallace MT, Vaughan JW, Stein BE, Norris JL, Schirillo JA (2003b) Visual localization ability influences cross-modal bias. J Cogn Neurosci 15:20–29.

Heron J, Whitaker D, McGraw F (2004) Sensory uncertainty governs the extent of audio-visual interaction. Vision Res 44:2875–2884.

Herrmann CS, Mecklinger A, Pfeifer E (1999) Gamma responses and ERPs in a visual classification task. Clin Neurophysiol 110:636–642.

Herrmann CS, Lenz D, Junge S, Busch NA, Maess B (2004) Memory-matches evoke human gamma-responses. BMC Neurosci 5:13.

Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ (2000) Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. Clin Neurophysiol 111:1745–1758.

Kaiser J, Hertrich I, Ackermann H, Lutzenberger W (2006) Gamma-band activity over early sensory areas predicts detection of changes in audiovisual speech stimuli. NeuroImage 30:1376–1382.

Lachaux JP, George N, Tallon-Baudry C, Martinerie J, Hugueville L, Minotti L, Kahane P, Renault B (2005) The many faces of the gamma band response to complex visual stimuli. NeuroImage 25:491–501.

Lehmann S, Murray MM (2005) The role of multisensory memories in unisensory object discrimination. Brain Res Cogn Brain Res 24:326–334.

Lenz D, Schadow J, Thaerig S, Busch NA, Herrmann CS (2007) What's that sound? Matches with auditory long-term memory induce gamma activity in human EEG. Int J Psychophysiol, in press.

Lutzenberger W, Pulvermüller F, Birbaumer N (1994) Words and pseudowords elicit distinct patterns of 30-Hz EEG responses in humans. Neurosci Lett 176:115–118.

Lutzenberger W, Pulvermüller F, Elbert T, Birbaumer N (1995) Visual stimulation alters local 40-Hz responses in humans: an EEG-study. Neurosci Lett 183:39–42.

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–748.

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cogn Brain Res 14:115–128.

Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. Cereb Cortex 14:452–465.

Müller MM, Bosch J, Elbert T, Kreiter A, Sosa MV, Sosa PV, Rockstroh B (1996) Visually induced gamma-based responses in human electroencephalographic activity: a link to animal studies. Exp Brain Res 112:96–102.

Murray MM, Wylie GR, Higgins BA, Javitt DC, Schroeder CE, Foxe JJ (2002) The spatiotemporal dynamics of illusory contour processing: combined high-density electrical mapping, source analysis, and functional magnetic resonance imaging. J Neurosci 22:5055–5073.

Murray MM, Foxe DM, Javitt DC, Foxe JJ (2004) Setting boundaries: brain dynamics of modal and amodal illusory shape completion in humans. J Neurosci 24:6898–6903.

Sams M, Aulanko R, Hamalainen M, Hari R, Lounasmaa OV, Lu ST, Simola J (1991) Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. Neurosci Lett 127:141–145.

Senkowski D, Talsma D, Herrmann CS, Woldorff MG (2005) Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. Exp Brain Res 166:411–426.

Senkowski D, Talsma D, Grigutsch M, Herrmann CS, Woldorff MG (2007) Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. Neuropsychologia 45:561–571.

Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, "unisensory: processing. Curr Opin Neurobiol 15:454–458.

Stekelenburg JJ, Vroomen J, de Gelder B (2004) Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. Neurosci Lett 357:163–166.

Tallon C, Bertrand O, Bouchet P, Pernier J (1995) Gamma-range activity evoked by coherent visual stimuli in humans. Eur J Neurosci 7:1285–1291.

Tallon-Baudry C (2003) Oscillatory synchrony and human visual cognition. J Physiol (Paris) 97:355–363.

Tallon-Baudry C, Bertrand O (1999) Oscillatory gamma activity in humans and its role in object representation. Trends Cogn Sci 3:151–162.

Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J (1996) Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. J Neurosci 16:4240–4249.

Tallon-Baudry C, Bertrand O, Delpuech C, Permier J (1997) Oscillatory gamma-band (30–70 Hz) activity induced by a visual search task in humans. J Neurosci 17:722–734.

Tallon-Baudry C, Bertrand O, Peronnet F, Pernier J (1998) Induced gamma-band activity during the delay of a visual short-term memory task in humans. J Neurosci 18:4244–4254.

Talsma D, Woldorff MG (2005) Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. J Cogn Neurosci 17:1098–1114.

Teder-Sälejärvi WA, McDonald JJ, Di Russo F, Hillyard SA (2002) An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. Brain Res Cogn Brain Res 14:106–114.

Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ (1997) Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. Proc Natl Acad Sci USA 94:14792–14797.

Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. Psychol Bull 88:638–667.

Witten IB, Knudsen EI (2005) Why seeing is believing: merging auditory and visual worlds. Neuron 48:489–496.