

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/143323>

Please be advised that this information was generated on 2021-10-17 and may be subject to change.

# The Rapid Extraction of Gist—Early Neural Correlates of High-level Visual Processing

Frank Oppermann<sup>1</sup>, Uwe Hassler<sup>2</sup>, Jörg D. Jescheniak<sup>1</sup>,  
and Thomas Gruber<sup>2</sup>

## Abstract

■ The human cognitive system is highly efficient in extracting information from our visual environment. This efficiency is based on acquired knowledge that guides our attention toward relevant events and promotes the recognition of individual objects as they appear in visual scenes. The experience-based representation of such knowledge contains not only information about the individual objects but also about relations between them, such as the typical context in which individual objects co-occur. The present EEG study aimed at exploring the availability of such relational knowledge in the time course of visual scene processing, using oscillatory evoked gamma-band responses as a neural correlate for a currently activated cortical stimulus representation. Participants decided whether

two simultaneously presented objects were conceptually coherent (e.g., mouse–cheese) or not (e.g., crown–mushroom). We obtained increased evoked gamma-band responses for coherent scenes compared with incoherent scenes beginning as early as 70 msec after stimulus onset within a distributed cortical network, including the right temporal, the right frontal, and the bilateral occipital cortex. This finding provides empirical evidence for the functional importance of evoked oscillatory activity in high-level vision beyond the visual cortex and, thus, gives new insights into the functional relevance of neuronal interactions. It also indicates the very early availability of experience-based knowledge that might be regarded as a fundamental mechanism for the rapid extraction of the gist of a scene. ■

## INTRODUCTION

When pictures are presented as briefly as 125 msec in rapid succession, we are nevertheless able to detect an object category reliably (Potter, 1975), demonstrating the outstanding speed of visual perception. Several studies have investigated how fast such object knowledge is activated using ERPs or intracranial recordings and have shown it to be available roughly 100–150 msec after stimulus onset (Liu, Agam, Madsen, & Kreiman, 2009; VanRullen & Thorpe, 2001; Thorpe, Fize, & Marlot, 1996). When scenes contain multiple objects, not only the knowledge related to individual objects comes into play but also the knowledge reflecting the general meaning of the scene, the so-called gist (Bar, 2004; Davenport & Potter, 2004; Hochstein & Ahissar, 2002; Henderson & Hollingworth, 1999). The extraction of gist from a visual scene is based on the conceptual coherence between its constituting elements. Although it has been demonstrated that physiological properties of the visual input, such as the spatial frequencies in the picture of a scene, contribute to the extraction of gist (e.g., Oliva & Torralba, 2001, 2007; Schyns & Oliva, 1994), the present study focused on the semantic relation between objects that fundamentally contributes to this mechanism.

The coherence between objects arises from several semantic scene properties, such as the probability of co-occurrence of objects in a common context or their relative position and size. Several studies have shown that the relational embedding of a target object in a scene facilitates its detection (e.g., Biederman, Mezzanotte, & Rabinowitz, 1982), whereas the embedding of the object in an unrelated scene (e.g., an octopus in a farm scene) hinders its detection. For example, objects that are not integrated in a scene are fixated more often and longer (Henderson, Weeks, & Hollingworth, 1999; Loftus & Mackworth, 1978) and attract more attention (Gordon, 2004) because they require a higher demand in processing. The coherence of objects also affects the processing of task-irrelevant context objects. For example, a naming study demonstrated that context objects in coherent scenes are more likely to be processed up to a lexical–phonological level (Oppermann, Jescheniak, & Schriefers, 2008; see also Oppermann, Jescheniak, Schriefers, & Gorges, 2010). Overall, these studies demonstrate that the gist provided by semantic relations of objects in a scene promotes visual processing.

The present study focused on (1) the temporal dynamics of gist extraction and (2) the cortical origins of gist perception. We hypothesized that the conceptual relations between multiple depicted objects are effective as early as incoming sensory information matches existing semantic memory traces in a first feed-forward sweep (Lamme &

<sup>1</sup>University of Leipzig, <sup>2</sup>University of Osnabrück

Roelfsema, 2000). To tap into the processing of these relations, we used high-density EEG (128 electrodes) and analyzed the so-called evoked gamma-band response (eGBR). eGBRs reflect cortical oscillatory activity in a frequency range above approximately 25 Hz. They occur at a latency range of approximately 100 msec after stimulus onset and are precisely phase-locked to stimulus onset (e.g., Tallon-Baudry & Bertrand, 1999). Cortical oscillatory activity in the gamma-band frequency range is supposed to reflect the activation of distributed cortical stimulus representations (Fries, 2005; Engel & Singer, 2001). This activation can be influenced by the existence of stimulus-specific memory representations (Roye, Schröger, Jacobsen, & Gruber, 2010; Herrmann, Lenz, Junge, Busch, & Maess, 2004; Herrmann, Munk, & Engel, 2004). For example, the study by Herrmann, Lenz, et al. (2004) demonstrated that real-world objects, which are represented in long-term memory, evoked a stronger eGBR at occipital electrodes than nonsense objects without a long-term memory representation. Hence, the authors argued that the eGBR reflects the integration of incoming sensory information and existing memory templates.

Our experiment contrasted stimulus displays, in which two objects shared a semantic–conceptual relation (coherent scene, e.g., mouse–cheese) with stimulus displays in which the two objects had no obvious relation (incoherent scene, e.g., crown–mushroom; see also Figure 1). We expected a stronger eGBR elicited by coherent scenes, reflecting the processing of relational information in addition to the processing of individual object information.

## METHODS

### Participants

Twenty-two healthy, right-handed adults participated in the experiment. All of them had normal or corrected-to-normal vision.

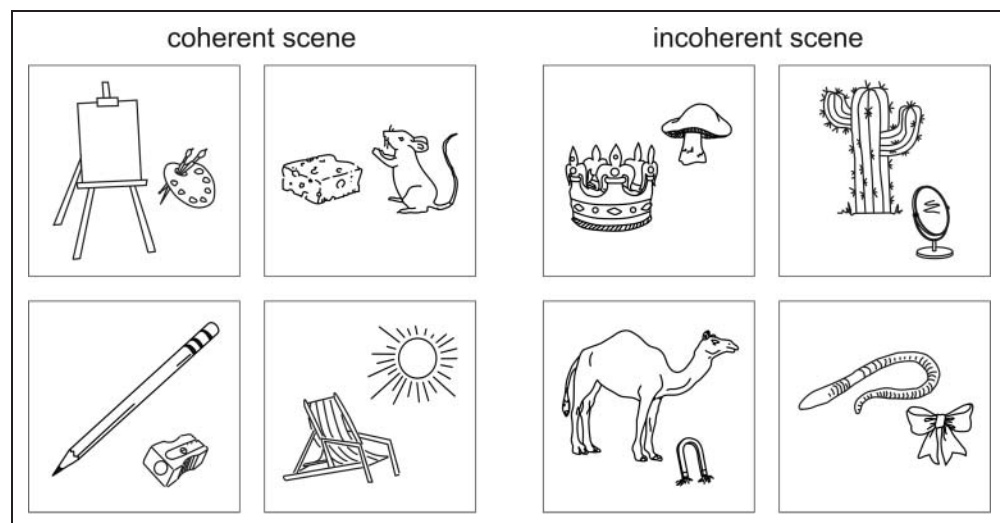
### Materials, Design, and Procedure

Three hundred twenty (320) line drawings of single objects were paired forming scenes consisting of two objects. Half the objects were combined to form a coherent context (coherent condition); the other half were arbitrarily combined (incoherent condition, 80 scenes per condition). Pictures of scenes were presented as white line drawings on a black screen (see Figure 1 for an example). Pictures of each scene were sized to fill an imaginary square of about  $8 \times 8$  cm, hence, were of same size in both conditions. The number of filled pixels was matched between conditions. Furthermore, we have analyzed the energy level across the spatial frequency range of our stimuli, because spatial frequency information of visual scenes may be attributed to the speed and different pathways of processing (e.g., Bar, 2003) and may convey different aspects of scene information (e.g., Schyns & Oliva, 1994). There were no significant differences between conditions when analyzing the complete frequency range, only low spatial frequency information (below four cycles per degree of visual angle) or only high spatial frequency information (above six cycles per degree of visual angle), all  $F_s < 1$  and no interaction with spatial frequency,  $F_s < 1$ .

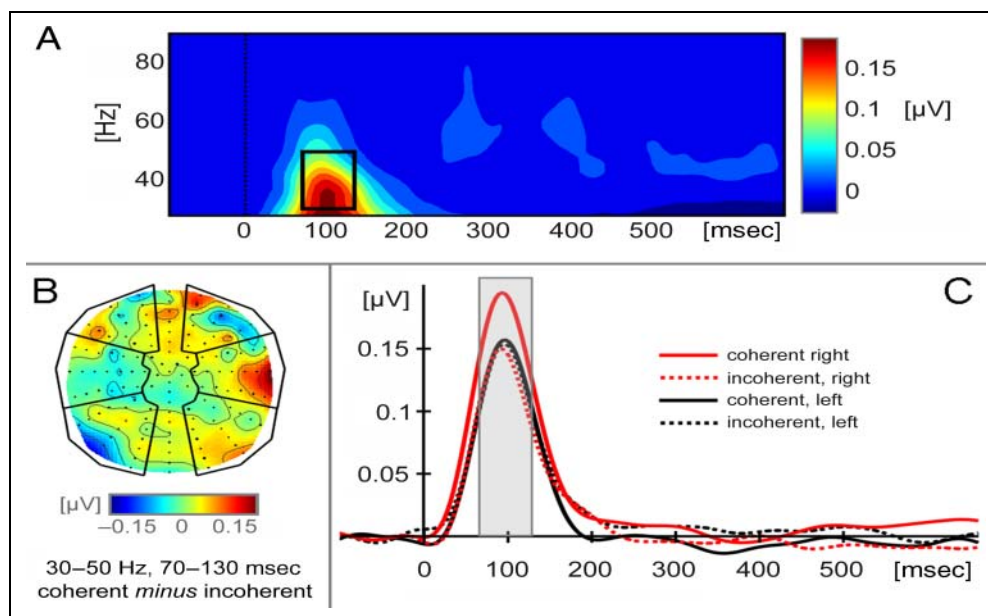
The sequence of the pictures was randomized for each participant. All objects were used only once, either in the coherent or in the incoherent condition. This procedure prevents participants from becoming familiar with any specific object and building up expectations with respect to the topic of any upcoming scene.

Participants were instructed to decide whether the object pairs shared a general meaning (coherent condition) or not (incoherent condition) by giving a push-button response as fast and as accurately as possible. The left-to-right assignment of response keys was counterbalanced across participants. Each trial lasted between 2900 and 3200 msec. First, a fixation cross appeared at the center of the screen placed 1.5 m in front of the participants (frame

**Figure 1.** Illustration of stimulus material. All pictures consisted of two objects. The four pictures on the left illustrate the coherent condition in which the objects in a scene were conceptually related, allowing for gist extraction. The four pictures on the right illustrate the incoherent condition, in which the objects were unrelated. The original displays in the experiment showed white line drawings on a black background.



**Figure 2.** eGBR. (A) The figure shows the grand mean baseline corrected TF plot of the stimulus-locked activity averaged across conditions and all 128 electrodes. eGBRs were elicited in the frequency window between 30 and 50 Hz in the time window of 70–130 msec (see black frame). (B) The difference map (coherent minus incoherent condition) shows the topography of the condition effect. The framed electrodes represent the six clusters that have been used for statistical analyses. The variable caudality compared the two anterior, central, and posterior electrode clusters. The variable hemisphere compared the three left with the three right electrode clusters. (C) Line plot of grand mean eGBRs illustrating the statistical interaction of condition (coherent vs. incoherent) and hemisphere (left vs. right). The gray bar marks the analyzed time window.



rate = 70 Hz). After a variable interval of 500–800 msec, the scene was shown for 700 msec (visual angle of approximately 5°). Picture onset was synchronized to the vertical retrace of the monitor. After the disappearance of the picture, the fixation cross remained on the screen for another 800 msec and was followed by a blank screen (900 msec).

### Data Acquisition

The EEG was recorded continuously from 128 active electrodes using a BioSemi Active Two amplifier system (sampling rate = 512 Hz). Eye movements and blinks were monitored by recording the horizontal and vertical EOG. Two additional electrodes (CMS = common mode sense and DRL = driven right leg; cf. [www.biosemi.com/faq/cms&drl.htm](http://www.biosemi.com/faq/cms&drl.htm)) were used as recording reference and ground. For further analysis, the average reference was used.

### Data Analysis—General

Artifact correction was performed by means of *statistical correction of artifacts in dense array studies* (Junghöfer, Elbert, Tucker, & Rockstroh, 2000). Furthermore, incorrect responses were excluded from further analyses (approximately 18 trials per participant). The average rejection rate after artifact correction was approximately 20% of the epochs. Next, the EEG was off-line averaged time-locked to stimulus onset (one epoch of 1500 msec included a 500-msec prestimulus baseline interval).

### Data Analysis—eGBR

Spectral changes in oscillatory activity were analyzed by means of Morlet wavelet transformations (Bertrand & Pantev, 1994) using a width of seven cycles per wavelet. This method provides a time-varying magnitude of the signal in each frequency band, leading to a Time  $\times$  Frequency (TF) representation of the data. To determine a suitable time and frequency window for the statistical analysis, a TF plot averaged across all conditions and all electrodes was used. Furthermore, TF plots of typical participants were generated to document the stability of eGBRs at a single-subject level. These analyses resulted in the selection of a TF window of 70–130 msec and 30–50 Hz. Next, we submitted the mean eGBR amplitudes (70–130 msec, 30–50 Hz) within six regional electrode means (see Figure 2B) to a repeated measure ANOVA involving the three fixed variables Coherence (coherent vs. incoherent object pairs), Caudality (anterior vs. central vs. posterior), and Hemisphere (left vs. right).

Finally, and to validate the oscillatory nature of the eGBR, we have band-pass filtered the ERP from 30 to 50 Hz and compared the topographies of successive peaks and troughs of the resulting signal in the time domain.<sup>1</sup>

### Source Analysis

To localize the cortical generators of the statistically significant eGBR differences between coherent and incoherent stimulus pairings, we applied variable resolution



electromagnetic tomography (VARETA; Bosch-Bayard et al., 2001). This procedure provides the spatial intracranial distribution of primary current densities (PCDs) in source space, which is best compatible with the amplitude distribution in electrode space. In particular, the eGBR was transformed into the frequency domain as described above (wavelet analysis), and VARETA was applied to the complex wavelet coefficients (cf. Gruber, Trujillo-Barreto, Giabbiconi, Valdes-Sosa, & Müller, 2006). Because of the linear relationship between EEG and PCD, the complex source reconstructions can be interpreted as an estimate of the wavelet coefficients of the PCD (complex inverse solution; Trujillo-Barreto, Aubert-Vazquez, & Valdes-Sosa, 2004). As possible sources of the signal, 3244 grid points (“voxels”) of a 3-D grid (7-mm grid spacing) were used. This grid and the arrangement of 128 electrodes were placed in registration with the average probabilistic MRI atlas (“average brain”) produced by the Montreal Neurological Institute (MNI; Evans et al., 1993). Statistical comparisons were carried out by means of Hottelling’s  $T^2$  tests to localize differences in activation between coherent and incoherent stimuli. Activation threshold corrections to account for spatial dependencies between voxels were calculated by means of random field theory (Worsley et al., 1996). Regarding all SPMs, the results were thresholded at a significance level of  $p < .01$ . Finally, the outcomes were depicted as 3-D activation images constructed on the basis of the MNI average brain.

### Data Analysis—ERPs

Before all ERP analyses, a 25-Hz low-pass filter was applied to the data. Furthermore, the averaged signal in a baseline period (500 msec prestimulus) was subtracted from all samples. To rule out perceptual differences in early visual processing between the two experimental conditions, we compared the following early ERP components: P1 (100–120 msec) and N1 (155–175 msec). Condition differences within these components were analyzed by means of paired  $t$  tests (averaged amplitudes across posterior electrodes). Additionally, we analyzed a late component (L1, 300–600 msec) at the six regional means presented in Figure 2B. Moreover, the ERP was examined within the same time window as the eGBR (70–130 msec), thereby using the same repeated measure ANOVA model as for the eGBR.

## RESULTS

### Behavioral Data

Behavioral data revealed faster RTs for coherent compared with incoherent object pairs (coherent: mean = 743 msec,  $SEM = 33$  msec; incoherent: mean = 846 msec,  $SEM = 35$  msec; mean difference = 103 msec,  $SEM = 8$  msec;  $t(21) = 13.91$ ;  $p < .0001$ ).

### eGBR

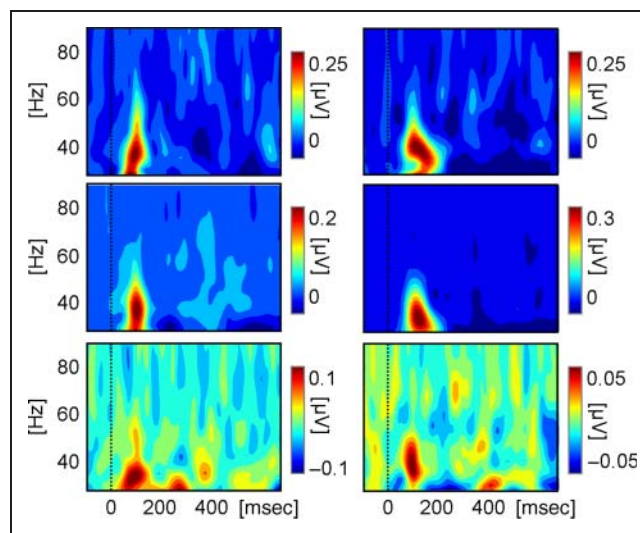
In Figure 2A, the TF plot of the stimulus-locked activity over all participants averaged across conditions and all electrodes is presented. Figure 3 shows TF plots of six typical participants. On the basis of these plots, an eGBR peak in a TF window of 30–50 Hz and 70–130 msec was selected for further analyses.<sup>2</sup>

We found an interaction of Coherence and Hemisphere ( $F(1, 21) = 6.91$ ,  $p < .02$ ,  $MSE = .303$ ), although no other effects involving the factor Coherence reached significance. Subsequent planned comparisons ( $t$  tests) revealed that the eGBR was significantly increased for coherent compared with incoherent object pairs in the right hemisphere (coherent: mean = 0.170  $\mu V$ ,  $SEM = 0.027$   $\mu V$ ; incoherent: mean = 0.128  $\mu V$ ,  $SEM = 0.025$   $\mu V$ ; mean difference = 0.042  $\mu V$ ,  $SEM = 0.016$   $\mu V$ ;  $t(21) = 2.66$ ,  $p < .02$ ), although no difference was observable in the left hemisphere (coherent: mean = 0.135  $\mu V$ ,  $SEM = 0.026$   $\mu V$ ; incoherent: mean = 0.133  $\mu V$ ,  $SEM = 0.025$   $\mu V$ ; mean difference = 0.003  $\mu V$ ,  $SEM = 0.012$   $\mu V$ ;  $t(21) < 1$ ; Figure 2C).

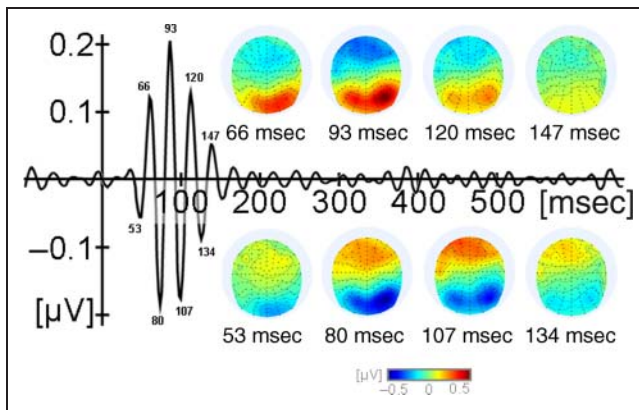
In Figure 4, the topographies of successive peaks and troughs of the band-pass filtered ERP from 30 to 50 Hz is presented. The result showed a stable topography across these peaks and troughs, indicating the oscillatory nature of the here reported eGBR effect.

### Source Analysis

The right hemispherical distribution of the coherence effect (Figure 2B) was supported by the VARETA that estimates the sources of the intracranial density distributions compatible with the observed scalp voltage topographies



**Figure 3.** Individual eGBR of six representative participants. The baseline corrected TF plots reveal different amplitudes and frequencies of the eGBR. The peaks of oscillatory activity vary from 30 to 50 Hz for individual participants.



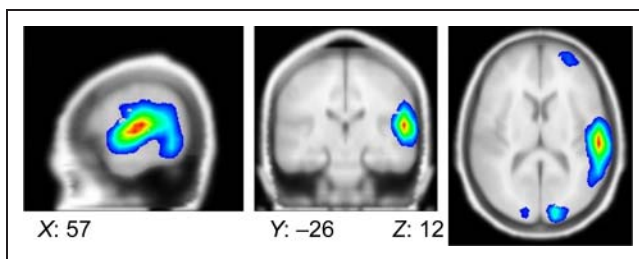
**Figure 4.** ERP at posterior electrode sites averaged across both conditions and band-pass filtered between 30 and 50 Hz (the evoked gamma-band range). The eight topographies reflect successive peaks and troughs averaged across both conditions at indicated time points. The peaks are presented above, and the troughs are below the graph.

(see also Gruber et al., 2006). This analysis revealed an activated network comprising the right middle frontal gyrus, the right middle and superior temporal gyrus, the precentral gyrus, and the bilateral occipital lobes as the source of the effect (Figure 5).

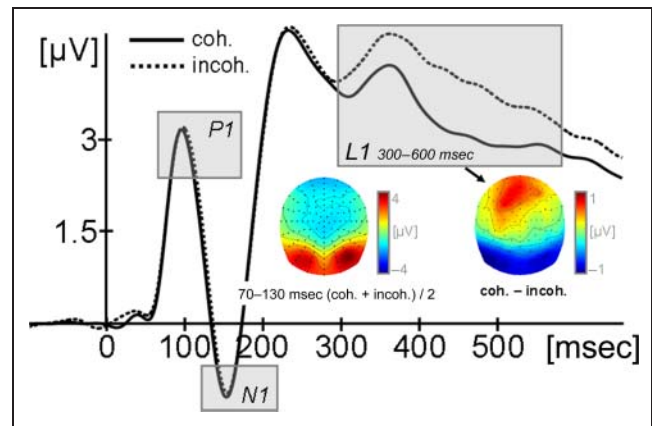
### ERPs

In the comparison of the coherent and incoherent condition at posterior electrodes, none of the amplitude differences of the early ERP components (P1, N1) reached significance (P1:  $t(21) < 1$ ; N1:  $t(21) = 1.20, p > .24$ ; see also Figure 6).

Regarding the L1 component (300–600 msec), the ANOVA resulted in a significant interaction of Coherence and Caudality ( $F(2, 42) = 83.84, p < .0001, MSE = .324$ ). Subsequent planned comparisons ( $t$  tests) revealed that the L1 component was significantly more negative for coherent compared with incoherent condition at posterior electrodes (see Figure 6; posterior coherent: mean =



**Figure 5.** Result of the VARETA. Statistically significant eGBR differences (coherent vs. incoherent;  $p < .01$ ) are depicted in color in sagittal, coronal, and axial slices (from left to right) containing the center of gravity of the observed effect.  $X, Y,$  and  $Z$  coordinates represent the location of the slices in MNI space. The hottest colors indicate the highest  $T^2$  values. The activated cortical network involves the right middle frontal gyrus and bilateral occipital areas (see the axial slice preferentially) as well as the right middle and superior temporal gyrus and the precentral gyrus (see the sagittal slice).



**Figure 6.** ERPs for the coherent (coh.) and incoherent (incoh.) condition averaged across posterior electrode sites (for electrodes see Figure 2B). The data represent the grand mean baseline corrected average across 22 participants. The left topography reflects the ERP in the time window of 70–130 msec averaged across both conditions. The right topography shows the difference of coherent minus incoherent condition in the L1 time window (300–600 msec).

4.07  $\mu\text{V}$ ,  $SEM = 0.65 \mu\text{V}$ ; posterior incoherent: mean = 5.37  $\mu\text{V}$ ,  $SEM = 0.65 \mu\text{V}$ ; mean difference = 1.30  $\mu\text{V}$ ,  $SEM = 0.12 \mu\text{V}$ ;  $t(21) = 10.89, p < .0001$ ), whereas the coherent condition was significantly more positive at central and anterior electrodes (central coherent: mean =  $-1.02 \mu\text{V}$ ,  $SEM = 0.18 \mu\text{V}$ ; central incoherent: mean =  $-1.35 \mu\text{V}$ ,  $SEM = 0.17 \mu\text{V}$ ; mean difference = 0.33  $\mu\text{V}$ ,  $SEM = 0.06 \mu\text{V}$ ;  $t(21) = 5.57, p < .0001$ ; anterior coherent: mean =  $-3.44 \mu\text{V}$ ,  $SEM = 0.68 \mu\text{V}$ ; anterior incoherent: mean =  $-4.25 \mu\text{V}$ ,  $SEM = 0.66 \mu\text{V}$ ; mean difference = 0.81  $\mu\text{V}$ ,  $SEM = 0.12 \mu\text{V}$ ;  $t(21) = 7.04, p < .0001$ ). This ANOVA result is underlined by the difference topography (coherent minus incoherent) in Figure 6 (right topography).

The ERP for the same regional means and the same latency range (70–130 msec) as the eGBR revealed no effects (in the ANOVA involving the three fixed variables Coherence, Caudality, and Hemisphere; none of the interactions involving the factor coherence nor the main effect of Coherence reached significance; all  $F < 2.3, p > .1$ ). Furthermore, Figure 6 (left topography) depicts the topography of the ERP averaged across both coherence conditions in the latency range from 70 to 130 msec to allow for a comparison with the topographies depicted in Figure 4.

In summary, our analyses show that the results regarding the eGBR are not accompanied by changes in early ERP components. They are specific to the gamma-band range. Thus, the transformation in the frequency domain reveals effects that are not visible in conventional ERP analyses.

### DISCUSSION

The present study investigated the temporal availability of semantic knowledge in the processing of visual scenes by using oscillatory activity in the gamma-band range as

a neural correlate of the formation of a coherent scene representation. We observed a very early effect in the eGBR (30–50 Hz) in a latency range of about 70–130 msec. Within this time window, the eGBR was increased for scenes in which objects were semantically related compared with unrelated objects. Thus, semantic scene knowledge is activated at early stages within the visual processing hierarchy. The results suggest that the difference of eGBRs between the coherent and the incoherent condition reflects the facilitated binding of two objects into a coherent scene representation.

However, before accepting this far-reaching conclusion, we have to rule out alternative explanations of this effect. Within the experiment, an individual object occurred only once, either in the coherent or in the incoherent condition. Because different objects were used in each condition, one might argue that, in fact, some physical differences between object sets used in both conditions may have caused the observed eGBR effect. Indeed, there are studies which have shown that eGBR measurements are sensitive to physical stimulus characteristics (e.g., Schadow et al., 2007; Busch, Debener, Kranczoch, Engel, & Herrmann, 2004; for a review, see Herrmann, Fründ, & Lenz, 2010). For example, a study by Busch et al. (2004) demonstrated that the eGBR increases with the size of a stimulus. Such variation in the eGBR is accompanied by concurrent changes in the early ERPs (P1 and N1; see also Martinovic, Gruber, & Müller, 2008; Busch, Herrmann, Müller, Lenz, & Gruber, 2006) and, most importantly, can be attributed to visual areas from the scalp topography. In our study, several reasons argue against the objection that physical stimulus characteristics between conditions might have caused the difference in eGBRs. In each condition, we used 160 different individual objects. To start with, two objects each were paired, forming object configurations that had the same size and contained an equal number of pixels in both conditions (see Materials). Furthermore, we observed no differences in early ERPs (P1, N1), which should have been affected by physical stimulus differences between picture sets. Finally, the eGBR effect in the present study has been localized to a network, including the right middle and superior temporal gyrus and the right middle frontal gyrus (see Figure 2B), whereas eGBR differences caused by physical stimulus properties exhibit occipital loci and show no lateralization. Taken together, all these arguments speak against the interpretation of our effect in terms of differences in physical stimulus properties. Furthermore, preliminary data from a study in which stimuli were counter-balanced across conditions by combining the same individual objects into coherent and incoherent pairs revealed a similar eGBR effect, as we observed in the present study (Oppermann, Hassler, & Gruber, in preparation). Taken all these points together, we can rule out that physical stimulus characteristics caused the observed effect. Rather, it has to be concluded that the semantic relation between objects is responsible for our finding.

A further point to address is the oscillatory nature of the observed effect in the gamma-band range. The TF plots of individual participants in Figure 3 show clear peaks of activity within the gamma-band range. This demonstrates that the effect is a genuine gamma-band effect and not only a residual of activity in a lower frequency range as suggested by the averaged TF plot in Figure 2A. This argument is further supported by the analyses of other frequency ranges in which no differences between coherent and incoherent conditions were found (see Footnote 2).

To corroborate the oscillatory nature of our eGBR effect, we have to demonstrate that the effect was not caused by successive peaks of activation of non-oscillatory events. This possibility was already discussed in previous studies for the auditory modality (Müller, Keil, Kissler, & Gruber, 2001; Pantev et al., 1993; Başar, Rosen, Başar-Eroglu, & Greitschus, 1987). According to this debate, the eGBR could reflect band-pass filtered portions of the middle-latency components. Thus, the eGBR would functionally equal an evoked response in the time domain. This alternative seems unlikely by the stability of the topographies across successive peaks and troughs of the band-pass filtered ERP in the gamma-band frequency range from 30 to 50 Hz (see Figure 4), because successive non-oscillatory events would be accompanied by changes of the topography. This finding supports the oscillatory nature of our eGBR effect.

Given all these arguments, our results demonstrate that early eGBRs reflecting cortical oscillatory activity are modulated by the semantic coherence of objects embedded in scenes. This finding provides first evidence that effects in eGBR measures are not confined to early visual areas reflecting the processing of physical stimulus characteristics or the top-down modulation of this processing by attention or stimulus anticipation (for a current review, see Herrmann et al., 2010). Rather, eGBRs might also reflect a first matching of visual input with stored long-term memory representations (see Herrmann, Munk, et al., 2004, for a related view) accompanied by the activation of a distributed network in areas beyond the visual cortex. Although the existence of visual eGBRs in structures beyond visual areas (with a central peak) was already demonstrated in previous studies, they were not shown to be sensitive to variations in stimulus type or other manipulations so far (e.g., Tallon-Baudry, Bertrand, Delpuech, & Pernier, 1997; Tallon-Baudry, Bertrand, Wienbruch, Ross, & Pantev, 1997). Thus, our data suggest that the presentation of multiple objects triggers the activation of a conjoint network, which represents the semantic coherence of objects. This conclusion supports the assumption that oscillatory activity of neuronal networks reflects the integration of cortically distributed representations into a coherent whole (Singer, 1999; Singer & Gray, 1995; von der Malsburg & Schneider, 1986). In the present case, this indicates the extraction of the gist of a visual scene. This binding by synchronization may reflect the preferred mechanism to bind individual object representations that can occur in



various constellations to a coherent scene representation (Singer, 1999). Whether this binding mechanism only becomes effective when the coherence between objects is task-relevant and attention is directed to this aspect or whether it is independent of the task relevance needs to be explored in further studies.

In recent years, there is growing evidence that visual processing can be affected by contextual memory at about or even before 100 msec after stimulus appearance (Chaumon, Drouet, & Tallon-Baudry, 2008; Meeren, Hadjikhani, Ahlfors, Hamalainen, & de Gelder, 2008; Pourtois, Rauss, Vuilleumier, & Schwartz, 2008). These effects were reflected in ERPs of the EEG or magnetic fields of the magneto-encephalogram and were located in the visual cortex. Why these processes appear in ERPs in the time domain whereas the coherence of scenes is indicated in oscillatory eGBRs needs to be addressed in future studies. However, it may be speculated that the effects in the ERPs are based on perceptual familiarity, which is stored in a kind of permanent memory representation (e.g., visual search in familiar and unfamiliar textures or upright and upside-down presented images). The individual objects of coherent and incoherent object configurations in the current study did not differ in their familiarity to the participants. Therefore, we did not expect any difference between conditions at this level. The difference between conditions in our study might be because of a more abstract level of scene perception at which the semantic relation between the objects is reflected. Considering the type of stimuli used (see Figure 1) and having realized that participants had not seen the configurations of the objects before, it is likely that the amount of activated neurons did not differ between both conditions at early sensory processing stages, which would have been reflected in ERP differences. Because the effect was present in the eGBRs, our data support the functional importance of oscillatory activity for the formation of coherent stimulus representations in the brain.

The source reconstruction of the present eGBR effect suggests that the extraction of scene information is achieved in a widely distributed network that involves temporal, precentral, frontal, and occipital cortical areas. The center of activity was localized to temporal areas of the right hemisphere. This finding is in line with studies showing that temporal areas are involved in the representation of global categorical and relational knowledge in humans as well as primates (e.g., Gronau, Neta, & Bar, 2008; Sugase, Yamane, Ueno, & Kawano, 1999). Furthermore, the spatio-temporal characteristic of our effect in temporal areas provides evidence for a processing that is mainly based on a feed-forward architecture in the ventral stream (e.g., Meeren et al., 2008; Serre, Oliva, & Poggio, 2007; Lamme & Roelfsema, 2000). That applies in particular to the nonrecurring presentation of our stimuli because, in contrast to previous studies (e.g., Liu et al., 2009; VanRullen & Thorpe, 2001; Thorpe et al., 1996), our participants could not develop any expectation of the

upcoming scene content that may have facilitated its processing in a top-down guided manner. In contrast, the involvement of occipital areas is known from previous studies (e.g., Herrmann, Lenz, et al., 2004) and is argued to reflect the integration of the incoming sensory information and the matched memory templates. Whether this integration into the network goes along a connection to temporal areas reflecting matched memory representations (Herrmann et al., 2010) or is forwarded by frontal areas (Strüber, Basar-Eroglu, Hoff, & Stadler, 2000; Tallon-Baudry, Bertrand, Delpuech, et al., 1997; Tallon-Baudry, Bertrand, Wienbruch, et al., 1997) cannot be decided based on our data. The involvement of the middle frontal gyrus in this network suggests a link to the attention network and might reflect the detection of a behaviorally relevant stimulus (Corbetta, Patel, & Shulman, 2008; Corbetta & Shulman, 2002). Otherwise, this frontal activation could also reflect a fast mechanism facilitating object recognition in a top-down guided manner (Bar et al., 2006; Bar, 2003). According to this mechanism, the partially analyzed global scene information is directly transmitted from the visual cortex to the pFC. On the basis of this global information, expectations are formed there about most likely interpretations of the visual input, which are then transmitted to the temporal cortex (Bar et al., 2006). In the temporal cortex, this top-down signal is integrated with the incoming bottom-up signal from the visual cortex. According to this assumption, the activation in the temporal cortex would not only reflect pure bottom-up activation in a feed-forward architecture but also the matching with a top-down process from frontal areas. Notably, all this activation is driven by the stimuli itself without any specific (semantic) expectation of the upcoming content. All mechanisms discussed here suggest a fast connectivity pattern of involved areas in our activated network.

What may be the reason of the seen hemispheric disparity? To speculate, the difference might stem from a general preference in analyzing global structural information in the right hemisphere and in analyzing local aspects in the left hemisphere (e.g., Malinowski, Hübner, Keil, & Gruber, 2002; Robertson & Ivry, 2000; Fink et al., 1996). This assumption is supported by a study suggesting that the right occipito-temporal cortex is involved in an initial and coarse processing of natural scenes whereas the left occipito-temporal cortex is involved in a more detailed processing of the parts of a scene (Peyrin et al., 2005). In the present study, participants directed their attention toward the general coherence of the scene that might be part of an initial and coarse processing, and thus, a right hemispheric source of our effect would be expected. Furthermore, the hemispheric disparity could also reflect that the effects are associated with the ventral attention network, which is largely lateralized to the right hemisphere (Corbetta & Shulman, 2002).

Finally, we want to address the late effect in the ERPs (300–600 msec). The observed L1 might be composed of



several subcomponents (e.g., P3a and P3b). Thus, the effect might reflect attentional orienting and stimulus evaluation processes. Moreover, the L1 could also reflect a motor-response-related shift in ERPs, given that the behavioral response is considerable faster in the coherent compared with the incoherent condition. However, detailed analyses of ERP components in this late time range are not motivated by the question of the present study, which focused on early effect of visual scene processing. Therefore, we refrain from a detailed interpretation of this component.

In summary, our data provide evidence for a powerful neuronal mechanism that may account for the integrated visual processing of scenes on the basis of the fast retrieval of semantic knowledge. The increase of cortical oscillatory activity here might indicate the extraction of the gist of a scene. It suggests the dynamic binding of neuronal assemblies, which establish a coherent high-level visual representation within no more than 100 msec.

## Acknowledgments

This work was supported by grants from the German Research Council. We thank S. Boigs for help in picture preparation and A. Roye for helpful discussions.

Reprint requests should be sent to Frank Oppermann, Department of Psychology, University of Leipzig, Seeburgstrasse 14-20, D-04103 Leipzig, Germany, or via e-mail: oppermann@uni-leipzig.de.

## Notes

1. We thank an anonymous reviewer for directing our attention to this point.
2. Analyses of other frequency ranges (5–15 and 15–25 Hz) in the time range around their peak of evoked oscillatory activity (5–15 Hz for the time range of 100–250 msec and 15–25 Hz for the time range of 50–150 msec) revealed no significant effects involving the factor coherence (no main effect of coherence and no interactions with the factor coherence in the ANOVA involving the three fixed variables Coherence, Caudality, and Hemisphere; all  $F < 1.6$ ,  $p > .2$ ).

## REFERENCES

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

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*, 617–629.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 449–454.

Başar, E., Rosen, B., Başar-Eroglu, C., & Greitschus, F. (1987). The associations between 40 Hz-EEG and the middle latency response of the auditory evoked potential. *International Journal of Neuroscience*, *33*, 103–117.

Bertrand, O., & Pantev, C. (1994). Stimulus frequency dependence of the transient oscillatory auditory evoked responses (40 Hz) studied by electric and magnetic recordings in human. In C. Pantev, T. Elbert, & B.

Lutkenhoner (Eds.), *Oscillatory event-related brain dynamics* (Vol. 271, pp. 231–242). New York: Plenum.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*, 143–177.

Bosch-Bayard, J., Valdes-Sosa, P., Virues-Alba, T., Aubert-Vazquez, E., John, E. R., Harmony, T., et al. (2001). 3D statistical parametric mapping of EEG source spectra by means of variable resolution electromagnetic tomography (VARETA). *Clinical Electroencephalography*, *32*, 47–61.

Busch, N. A., Debener, S., Kranczioch, C., Engel, A. K., & Herrmann, C. S. (2004). Size matters: Effects of stimulus size, duration and eccentricity on the visual gamma-band response. *Clinical Neurophysiology*, *115*, 1810–1820.

Busch, N. A., Herrmann, C. S., Müller, M. M., Lenz, D., & Gruber, T. (2006). A cross-laboratory study of event-related gamma activity in a standard object recognition paradigm. *Neuroimage*, *33*, 1169–1177.

Chaumon, M., Drouet, V., & Tallon-Baudry, C. (2008). Unconscious associative memory affects visual processing before 100 ms. *Journal of Vision*, *8*, 10.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, *58*, 306–324.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*, 215–229.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*, 559–564.

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, *5*, 16–25.

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., & Peters, T. M. (1993). 3D statistical neuroanatomical models from 305 MRI volumes. In L. A. Klaisner (Ed.), *Nuclear science symposium and medical imaging conference* (Vol. 1–3, pp. 1813–1817). New York: IEEE.

Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1996). Where in the brain does visual attention select the forest and the trees? *Nature*, *382*, 626–628.

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, *9*, 474–480.

Gordon, R. D. (2004). Attentional allocation during the perception of scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 760–777.

Gronau, N., Neta, M., & Bar, M. (2008). Integrated contextual representation for objects' identities and their locations. *Journal of Cognitive Neuroscience*, *20*, 371–388.

Gruber, T., Trujillo-Barreto, N. J., Giabbiconi, C. M., Valdes-Sosa, P. A., & Müller, M. M. (2006). Brain electrical tomography (BET) analysis of induced gamma-band responses during a simple object recognition task. *Neuroimage*, *29*, 888–900.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271.

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 210–228.

Herrmann, C. S., Fründ, I., & Lenz, D. (2010). Human gamma-band activity: A review on cognitive and behavioral correlates and network models. *Neuroscience and Biobehavioral Reviews*, *34*, 981–992.

Herrmann, C. S., Lenz, D., Junge, S., Busch, N. A., & Maess, B. (2004). Memory-matches evoke human gamma-responses. *BMC Neuroscience*, *5*, 13.

- Herrmann, C. S., Munk, M. H. J., & Engel, A. K. (2004). Cognitive functions of gamma-band activity: Memory match and utilization. *Trends in Cognitive Sciences*, 8, 347–355.
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36, 791–804.
- Jungthöfer, M., Elbert, T., Tucker, D. M., & Rockstroh, B. (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, 37, 523–532.
- Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23, 571–579.
- Liu, H. S., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron*, 62, 281–290.
- Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 565–572.
- Malinowski, P., Hübner, R., Keil, A., & Gruber, T. (2002). The influence of response competition on cerebral asymmetries for processing hierarchical stimuli revealed by ERP recordings. *Experimental Brain Research*, 144, 136–139.
- Martinovic, J., Gruber, T., & Müller, M. M. (2008). Coding of visual object features and feature conjunctions in the human brain. *Plos One*, 3, e3781.
- Meeren, H. K. M., Hadjikhani, N., Ahlfors, S. P., Hamalainen, M. S., & de Gelder, B. (2008). Early category-specific cortical activation revealed by visual stimulus inversion. *Plos One*, 3, e3503.
- Müller, M. M., Keil, A., Kissler, J., & Gruber, T. (2001). Suppression of the auditory middle-latency response and evoked gamma-band response in a paired-click paradigm. *Experimental Brain Research*, 136, 474–479.
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42, 145–175.
- Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, 11, 520–527.
- Oppermann, F., Hassler, U., & Gruber, T. (in preparation). Effects of high-level scene processing on the evoked gamma-band response are task dependent.
- Oppermann, F., Jescheniak, J. D., & Schriefers, H. (2008). Conceptual coherence affects phonological activation of context objects during object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 587–601.
- Oppermann, F., Jescheniak, J. D., Schriefers, H., & Görges, F. (2010). Semantic relatedness among objects promotes the activation of multiple phonological codes during object naming. *Quarterly Journal of Experimental Psychology*, 63, 356–370.
- Pantev, C., Elbert, T., Makeig, S., Hampson, S., Eulitz, C., & Hoke, M. (1993). Relationship of transient and steady-state auditory evoked fields. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 88, 389–396.
- Peyrin, C., Schwartz, S., Seghier, M., Michel, C., Landis, T., & Vuilleumier, P. (2005). Hemispheric specialization of human inferior temporal cortex during coarse-to-fine and fine-to-coarse analysis of natural visual scenes. *Neuroimage*, 28, 464–473.
- Potter, M. C. (1975). Meaning in visual search. *Science*, 187, 965–966.
- Pourtois, G., Rauss, K. S., Vuilleumier, P., & Schwartz, S. (2008). Effects of perceptual learning on primary visual cortex activity in humans. *Vision Research*, 48, 55–62.
- Robertson, L. C., & Ivry, R. (2000). Hemispheric asymmetries: Attention to visual and auditory primitives. *Current Directions in Psychological Science*, 9, 59–63.
- Roye, A., Schröger, E., Jacobsen, T., & Gruber, T. (2010). Is my mobile ringing? Evidence for rapid processing of a personally significant sound in humans. *Journal of Neuroscience*, 30, 7310–7313.
- Schadow, J., Lenz, D., Thaeerig, S., Busch, N. A., Fründ, I., Rieger, J. W., et al. (2007). Stimulus intensity affects early sensory processing: Visual contrast modulates evoked gamma-band activity in human EEG. *International Journal of Psychophysiology*, 66, 28–36.
- Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges—Evidence for time-scale-dependent and spatial-scale-dependent scene recognition. *Psychological Science*, 5, 195–200.
- Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, U.S.A.*, 104, 6424–6429.
- Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, 24, 49–65.
- Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, 18, 555–586.
- Strüber, D., Basar-Eroglu, C., Hoff, E., & Stadler, M. (2000). Reversal-rate dependent differences in the EEG gamma-band during multistable visual perception. *International Journal of Psychophysiology*, 38, 243–252.
- Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, 400, 869–873.
- Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, 3, 151–162.
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1997). Oscillatory gamma-band (30–70 Hz) activity induced by a visual search task in humans. *Journal of Neuroscience*, 17, 722–734.
- Tallon-Baudry, C., Bertrand, O., Wienbruch, C., Ross, B., & Pantev, C. (1997). Combined EEG and MEG recordings of visual 40 Hz responses to illusory triangles in human. *NeuroReport*, 8, 1103–1107.
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520–522.
- Trujillo-Barreto, N. J., Aubert-Vazquez, E., & Valdes-Sosa, P. A. (2004). Bayesian model averaging in EEG/MEG imaging. *Neuroimage*, 21, 1300–1319.
- VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, 13, 454–461.
- von der Malsburg, C., & Schneider, W. (1986). A neural cocktail-party processor. *Biological Cybernetics*, 54, 29–40.
- Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4, 58–73.