

Gamma-band activity reflects multisensory matching in working memory

Daniel Senkowski · Till R. Schneider ·
Frithjof Tandler · Andreas K. Engel

Received: 27 October 2008 / Accepted: 28 April 2009 / Published online: 21 May 2009
© Springer-Verlag 2009

Abstract In real-world situations, the integration of sensory information in working memory (WM) is an important mechanism for the recognition of objects. Studies in single sensory modalities show that object recognition is facilitated if bottom-up inputs match a template held in WM, and that this effect may be linked to enhanced synchronization of neurons in the gamma-band (>30 Hz). Natural objects, however, frequently provide inputs to multiple sensory modalities. In this EEG study, we examined the integration of semantically matching or non-matching visual and auditory inputs using a delayed visual-to-auditory object-matching paradigm. In the event-related potentials (ERPs) triggered by auditory inputs, effects of semantic matching were observed after 120–170 ms at frontal and posterior regions, indicating WM-specific processing across modalities, and after 250–400 ms over medial-central regions, possibly reflecting the contextual integration of sensory inputs. Additionally, total gamma-band activity (GBA) with medial-central topography after 120–180 ms was larger for matching compared to non-matching trials. This demonstrates that multisensory matching in WM is reflected by GBA and that dynamic coupling of neural populations in this frequency range might be a crucial mechanism for integrative multisensory processes.

Keywords Multisensory · Crossmodal · Working memory · Gamma · EEG

Daniel Senkowski and Till R. Schneider contributed equally to the manuscript.

D. Senkowski (✉) · T. R. Schneider · F. Tandler · A. K. Engel
Department of Neurophysiology and Pathophysiology,
University Medical Center Hamburg-Eppendorf,
Martinistr. 52, 20246 Hamburg, Germany
e-mail: dsenkows@uke.uni-hamburg.de

Introduction

Objects in our environment frequently generate inputs to multiple sensory modalities. These inputs can enter our sensory system with a temporal asynchrony, like an approaching train, which we first see in the distance before we hear it. How efficiently such sensory inputs are integrated critically depends on the semantic matching with the template held in WM. As shown for the processing *within* single modalities, the recognition of inputs that match the WM contents is normally facilitated compared to inputs, which do not match WM contents (Elliott and Dolan 1999). Thus far, it is not well-understood how sensory inputs are integrated in WM *across* modalities and how this integration may be affected by semantic matching.

Indications about how semantic matching in WM could influence the integration of sensory inputs may be derived from implicit multisensory priming paradigms (Orgs et al. 2006; Schneider et al. 2008a, 2008b). In these paradigms a prime stimulus in one modality, e.g., a picture of an object, is followed by a target stimulus in a second modality, e.g., the sound of an object. The prime stimulus, which is not task-relevant, is either semantically congruent or incongruent with the target. Prime and target stimuli that conceptually represent the same object are defined as *matching*, and prime-target pairs that conceptually represent a different object are labeled *non-matching*. In this setup, reaction times (RTs) are typically shorter for matching compared to non-matching prime-target pairs. In the human electroencephalogram (EEG), the effect of semantic matching is often expressed by modulations of the N400 component in the event-related potentials (ERP) (Orgs et al. 2006; Schneider et al. 2008b), which most likely reflects contextual integrative processing (Kutas and Hillyard 1980a). Interestingly, a more recent EEG study has demonstrated

that multisensory semantic matching is also linked to gamma-band activity (GBA) (Schneider et al. 2008b). This finding is of particular note since an increasing number of studies suggest that temporal synchronization of neural activity may be an important mechanism for the integration of object features across modalities (Senkowski et al. 2006; Lakatos et al. 2007; Maier et al. 2008; for a recent review see Senkowski et al. 2008).

In line with these reports, human (Tallon-Baudry et al. 1998; Siegel et al. 2007) and animal studies (Womelsdorf et al. 2007; Engel et al. 2001) within single modalities provide strong evidence that the coupling of neural populations by temporal synchronization, in particular in the gamma-band (>30 Hz), may be crucial for the integration of object features in cortical networks (for review, see Engel et al. 2001). Additionally, neural synchronization mechanisms in the gamma-band are associated with WM processes in delayed matching-to-sample paradigms during the retention (Tallon-Baudry et al. 1998; Jokisch and Jensen 2007) and retrieval (Lutzenberger et al. 2002) of information. Herrmann and colleagues (2004b) have, therefore, proposed a match-and-utilization model (MUM), which postulates that GBA reflects the matching of incoming bottom-up information with a template held in WM. A central prediction of this model is that matching of bottom-up inputs with a template held in WM is paralleled by an increase in GBA. Together, these studies show an involvement of temporal synchronization mechanisms in integrative object feature processing and WM functions. Therefore, it can be predicted that differences in the integration of semantically matching and non-matching information across modalities should be also reflected in differences in GBA. Specifically, a larger GBA for stimuli that are semantically matching compared to inputs that are non-matching across modalities is predicted. Additionally, effects of semantic matching are expected for the N400 component of the ERP that has been related to the integration of contextual information, and which is particularly pronounced for non-matching information (Kutas and Hillyard 1980b). Finally, effects of semantic matching in the ERP may also be found at shorter latencies, which would fit with previous studies on the integrative processing of semantic audiovisual inputs (Senkowski et al. 2007a; Stekelenburg and Vroomen 2007)

To test these hypotheses, a paradigm derived from the one recently introduced by Schneider et al. (2008b) was used. In Schneider et al., participants were presented with task-irrelevant pictures of naturalistic objects prior to the presentation of task-relevant sounds of objects from either the same or a different object. The task was to indicate whether the object from which the sound was presented fits into a shoebox or not. Priming effects of semantic matching were found for the early GBA (around 150 ms) and for the N400 component. Importantly, no task was required in

response to the visual inputs. Thus, the semantic priming effects on auditory targets in this study were *implicit*. In contrast, in the present study, participants were instructed to attend to the visual and auditory inputs and to indicate whether auditory and visual stimuli belong to the same or to a different object. This task *explicitly* required the semantic matching of sensory inputs across modalities in WM. As a main result, effects of semantic matching were observed for GBA and in the ERP starting around 120 ms after sound onset.

Methods

Participants

Twenty-three healthy volunteers participated in the experiment (11 females, mean age 25.2, range 21–33 years) and received monetary compensation for their participation. All participants were native German speakers, had normal hearing (hearing loss <30 dB); normal or corrected to normal vision (visus >0.9); and reported no history of neurological or psychiatric illness. The data of four subjects with less than 32% remaining trials (i.e., less than 54 trials) per condition after artifact rejection in the EEG excluded from the further analysis. The study was conducted in accordance with the Declaration of Helsinki and informed consent was obtained from all participants prior to the recordings.

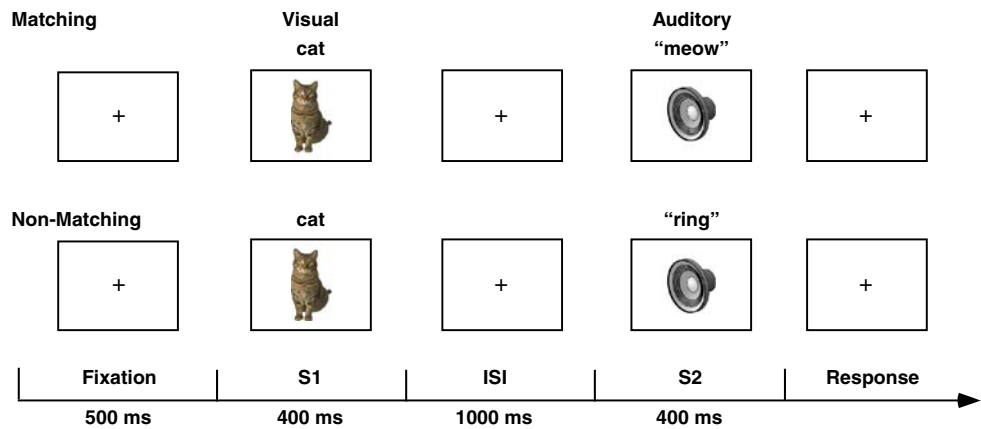
Stimuli

Visual stimuli were taken from a pool of 320 color photographs of natural objects for which norming data were available (Schneider et al. 2008a). The stimuli were presented centrally for 400 ms, subtending a visual angle of 8.5° vertically and 9° horizontally. Auditory stimuli were taken from a pool of 270 environmental sounds of natural objects from the same norming study. The intensities of all sounds were adjusted by equalizing the root mean square power across all sound files. To avoid on- and offset clicking transients the sound files were windowed with a linear 10-ms rise and fall time. Sounds (22 kHz, 16-bit, mono) were presented for 400 ms via Eartone foam-protected air tube earphones (AeroCompany, Indianapolis, IN) at 70 dB SPL. The presentation of stimuli was controlled using Presentation 0.80 (Neurobehavioral Systems, San Francisco, CA).

Procedure

Each trial consisted of a picture of an object (i.e., S1) and a sound of an object (i.e., S2) presented successively with a stimulus onset asynchrony (SOA) of 1,400 ms (Fig. 1).

Fig. 1 Experimental setup. Participants were instructed to indicate after the sound (S2) presentation whether the visual-auditory (S1-S2) stimulus pairs were either semantically matching or non-matching, representing conceptually the same or different objects. Thus, the task explicitly required the semantic matching of stimuli from different sensory modalities in working memory



Auditory-visual stimulus pairs were either semantically matching (50%) or semantically non-matching (50%), representing conceptually the same or different objects. The matching ($n = 85$) and non-matching stimulus sets ($n = 85$) were balanced for familiarity, naming accuracy, and correct object categorization according to the previously obtained stimulus norms (Schneider et al. 2008a). Overall, 340 trials (170 per condition) were presented in the experiment.

None of the auditory stimuli were used in both the matching and non-matching stimulus set. Additionally, it was verified that, on average, there was no difference between the time–frequency profiles of the sounds in the two stimulus sets. To this end, a time–frequency analysis (100 Hz–10 kHz) on the individual sound files with a frequency resolution of 100 Hz and a time resolution of 5 ms was performed. Running t tests for differences of the average power values in each time–frequency bin showed significant differences ($p < 0.05$) in 389 of 8,800 bins, i.e., in less than 5% of the bins, which would be expected by chance.

Stimuli were presented in a pseudo-randomized order across both conditions (matching, non-matching). Participants were instructed to indicate as accurately and quickly as possible after the presentation of the auditory input (S2) whether the two stimuli presented to both modalities were representing the same or a different object. Thus, the task combined the need for object identification and object matching with a speeded behavioral response, which is a prerequisite for reaction time (RT) analysis. Responses were given by pressing one of two buttons with the left or the right thumb, counterbalanced across participants. If no response was given within a time window of 2,000 ms, a message was presented on the screen reminding the participants to respond faster. Violation of the time limit occurred in less than 3% of the presented trials.

Data acquisition

EEG data were collected from 126 scalp sites using sintered Ag/AgCl ring electrodes mounted on an elastic cap (Falk

Minow Services, Herrsching, Germany). During recordings, the nose tip was used as reference, but prior to analysis the data were re-referenced to common average. Two additional electrodes were positioned below the eyes to record the electrooculogram. The data were recorded with an analog passband of 0.016–250 Hz and digitized at a sampling rate of 1,000 Hz using BrainAmp amplifiers (BrainProducts, Munich, Germany). Electrode impedances were kept below 20 k Ω (e.g., Ferree et al. 2001). Analysis of behavioral and EEG data was performed using Matlab 7.4 (MathWorks, Natick, MA) together with EEGLAB (Delorme and Makeig (2004); <http://www.sccn.ucsd.edu/eeGLAB>) and Fieldtrip (<http://www.ru.nl/fcdonders/fieldtrip>), two freely available open source toolboxes for EEG data analysis. For all further analyses the EEG data were band-pass filtered (0.3–110 Hz) and downsampled to 250 Hz. For the artifact removal a similar procedure as previously used was applied (Debener et al. 2005). First, epochs containing non-stereotyped artifacts (e.g., cable movement, swallowing) were manually removed. Then, extended infomax independent component analysis (ICA) was applied, using a weight change $< 10^{-7}$ as stop criterion. Independent components representing artifacts such as eyeblinks, horizontal eye movements, or electrocardiographic activity were removed from the EEG data by back-projecting all but these components. To remove muscular artifacts, trials were automatically rejected exceeding an amplitude criterion of $\pm 100 \mu\text{V}$ or a frequency criterion of 25 dB in the 20–40 Hz power spectrum. The number of trials rejected due to artifacts and due to incorrect responses varied across subjects. On average 54.5% of trials remained in the analysis.

Analysis of behavioral data

Reaction times (RTs) and error rates were compared directly between semantically matching and semantically non-matching trials using repeated measures analysis of variance (ANOVA).

Analysis of event-related potentials

Epochs for the analysis of ERPs were extracted starting 200 ms before S2 onset and lasting 1,200 ms. The interval from -200 ms to stimulus onset served as baseline. A 30-Hz lowpass filter was applied for the analysis of ERPs. An exploratory analysis of the ERP responses between the responses to visual (S1) stimuli for semantically matching and semantically non-matching trials revealed no significant differences. Therefore, S1 stimuli were not submitted to the further analysis. Only trials with correct behavioral responses were entered to the analysis of ERPs. In line with our previous study (Schneider et al. 2008b), nine regions of interest (ROI) were defined for the statistical analysis of ERPs. Each ROI comprised the averaged signal of six adjacent electrodes. Regions were defined as left frontal (LF), middle frontal (MF), right frontal (RF), left central (LC), middle central (MC), right central (RC), left posterior (LP), middle posterior (MP), and right posterior (RP). Statistical analysis was performed in two steps. First, to obtain information about the temporal evolution of significant differences, running t tests were performed on the ERPs investigating differences between responses to semantically matching and non-matching trials for each electrode and at each sampling point. An interval was considered to differ significantly between the conditions if at least five consecutive data points (i.e., 20 ms) reached a P value below 0.05 (Guthrie and Buchwald 1991). Second, a repeated measure ANOVA was performed using the within-subject factors Matching (matching, non-matching) and ROI (9 regions) for a short and a longer latency response interval. The short latency interval ranged from 120 to 170 ms and the longer latency interval from 250 to 400 ms.

Analysis of gamma-band activity

Spectral changes in neural synchronization were analyzed using a wavelet transform, which provides a good compromise between time and frequency resolution (Tallon-Baudry et al. 1998). Time–frequency representations were computed using a wavelet transformation for each channel by convolving the data with a complex Morlet wavelet $w(t, f_0)$ with a constant $Q = f_0/\sigma_f = 7$ for frequencies from 20 to 100 Hz (step size 1 Hz). Before averaging, all frequency transformations were performed at the single-trial level. Thus, the resulting total power contains signal components phase-locked and non-phase-locked to the stimulus. The resulting power was baseline-corrected for each frequency to obtain the relative signal change: $P(t, f)_{\text{corrected}} = 100 \times (P(t, f) - P_{\text{baseline}}(f))/P_{\text{baseline}}(f)$. The S2 prestimulus period (-300 to -100 ms) served as baseline for all spectral analyses. Only trials with correct behavioral responses were entered into the analysis of GBA. Grand mean time-fre-

quency representations were computed over all participants. Additionally, evoked power was calculated with the same parameters as for the total power analysis on the averaged signals. A repeated measure ANOVA with the factors Matching (matching, non-matching) and ROI (9 regions, identical to the ROIs used for the ERP analysis) was used to statistically analyze the spectral activity in the total and evoked gamma-band activity. In line with previous reports, the analysis of evoked GBA was conducted for a short latency interval between 50 and 100 ms (Senkowski et al. 2005, 2007b), whereas the analysis of total GBA was computed for an interval between 120 and 180 ms (Schneider et al. 2008b).

Results

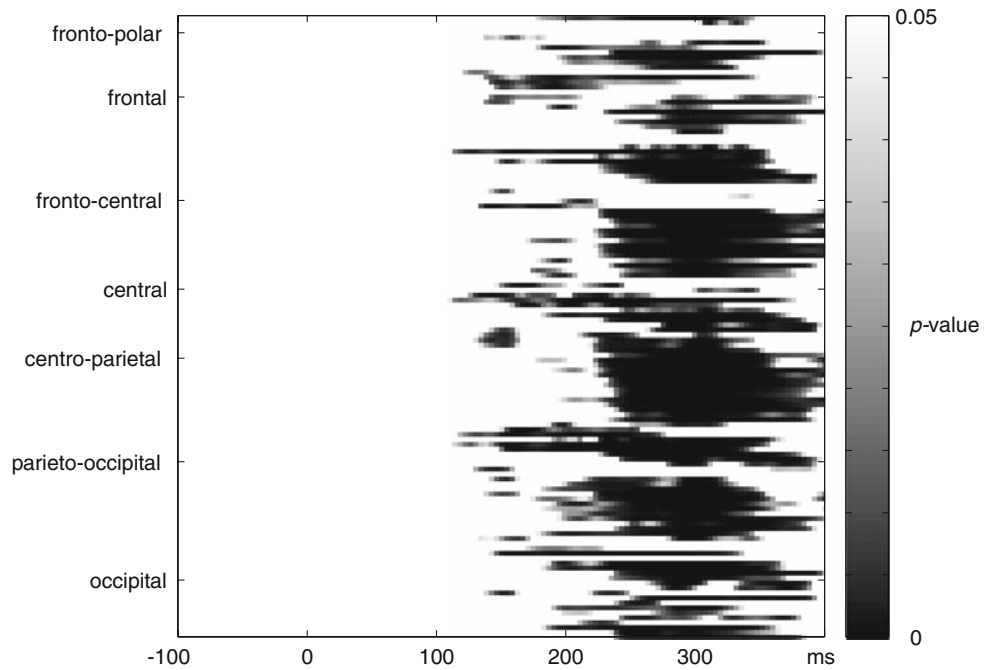
Behavioral data

The repeated measure ANOVA for RTs of semantically matching (711 ms) and semantically non-matching (719 ms) S2 stimuli did not reveal significant differences ($F < 1$). However, for the analysis of error rates significant differences between matching and non-matching trials were observed ($F_{1,18} = 65.2, p < 0.0001$). A higher error rate was found for matching (22.1%) compared to non-matching (6.2%) trials. Since the processing of matching trials was more difficult than the processing of non-matching trials, it may be that differences in difficulty have masked possible effects of semantic matching on RTs. To examine this possible masking effect, an analysis of covariances (ANCOVA) was computed between semantically matching and non-matching trials using RTs as dependent variable and the algebraic differences in error rate between matching and non-matching trials as covariate. The ANCOVA between matching and non-matching trials revealed significant differences in RTs ($F_{1,17} = 5.7, p < 0.05$), suggesting that the possible influence of semantic matching on RTs may have been masked by differences in processing difficulty between the two stimulus types.

Event-related potentials

The analysis of ERPs to the auditory S2 stimuli revealed two major time intervals of effects between semantic matching and non-matching trials. Figure 2 shows the results of the point-wise running t test for 124 scalp electrodes comparing matching versus non-matching trials. There were no effects before 120 ms following sound onset. For the 120–170 ms time interval, however, significant effects were found at frontal and posterior electrode sites. For the longer latency time interval (250–400 ms), significant integration effects were observed at multiple locations across the scalp.

Fig. 2 Point-wise running t tests between ERPs to matching and non-matching trials for 124 scalp electrodes. The earliest effects were found around 120–170 ms after the onset of auditory S2 inputs. Moreover, temporally sustained effects were observed starting at about 250 ms after auditory stimulus onset



The ANOVA for the interval between 120 and 170 ms revealed a main effect of the factor Matching ($F_{1,18} = 14.07$, $p < 0.0001$) and a significant Matching \times ROI interaction ($F_{1,18} = 4.53$, $p < 0.001$) (Fig. 3). Follow-up ANOVAs for the nine ROIs separately using the factor Matching revealed significant effects at the following ROIs: Right-frontal (matching $-1.57 \mu\text{V}$, non-matching $-1.09 \mu\text{V}$; $F_{1,18} = 14.2$, $p < 0.01$), left-frontal (matching $-1.02 \mu\text{V}$, non-matching $-0.69 \mu\text{V}$; $F_{1,18} = 5.37$, $p < 0.05$), left-posterior (matching $1.33 \mu\text{V}$, non-matching $0.88 \mu\text{V}$; $F_{1,18} = 8.88$, $p < 0.01$), and middle-posterior (matching $2.14 \mu\text{V}$, non-matching $1.78 \mu\text{V}$; $F_{1,18} = 6.93$, $p < 0.05$). No significant effects were found for the other ROIs. For the longer latency interval between 250 and 400 ms main effects were observed for the factor Matching ($F_{1,18} = 25.49$, $p < 0.0001$), ROI ($F_{1,18} = 39.72$, $p < 0.0001$), and the Matching \times ROI interaction ($F_{1,18} = 8.05$, $p < 0.0001$). A significant smaller deflection was found for non-matching ($-0.18 \mu\text{V}$) than for matching trials ($0.195 \mu\text{V}$) at the left-central ROI ($F_{1,18} = 5.3$, $p < 0.05$), at the middle-central ROI (non-matching $-1.77 \mu\text{V}$, matching $-0.17 \mu\text{V}$, $F_{1,18} = 31.96$, $p < 0.0001$), and at the right-central ROI (non-matching $0.065 \mu\text{V}$, matching $0.435 \mu\text{V}$, $F_{1,18} = 6.73$, $p < 0.05$). The mean differences and standard errors of the mean between the two conditions are provided in Table 1.

The topographical distribution of the ERPs (Fig. 3b) for the two latency intervals shows distinct patterns, which supports the assumption that different underlying neural generators are responsible for the short and long latency effects. Short latency effects (120–170 ms) were observed at frontal and posterior sites, whereas longer -latency effects

Table 1 Means and standard errors of the mean for the differences in the ERP and total GBA

	<i>M</i>	<i>SEM</i>
ERP 120–170 ms (μV , matching minus non-matching)		
Left frontal	0.332*	0.143
Right frontal	0.477**	0.127
Left posterior	-0.447**	0.15
Middle posterior	-0.362*	0.137
ERP 250–400 ms (μV , matching minus non-matching)		
Left central	-0.375*	0.163
Middle central	-1.598***	0.283
Right central	-0.37*	0.143
Total GBA 120–180 ms (% change, matching minus non-matching)		
Left frontal	6.143**	1.927
Middle central	9.782**	2.635

M mean, *SEM* standard error of the mean

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(250–400 ms) were found at medial-central, as well as left- and right-central regions.

Gamma-band activity

Figure 4 shows the time–frequency representation of the total GBA (20–100 Hz) for the medial-central ROI. The sound onset is followed by an increase in GBA starting around 50 ms. For the total GBA (30–40 Hz) the ANOVA in the 120–180 ms time window revealed a significant effect of Matching ($F_{1,18} = 4.79$, $p < 0.05$), and a Matching \times ROI interaction ($F_{1,18} = 2.96$, $p < 0.01$), showing enhanced GBA

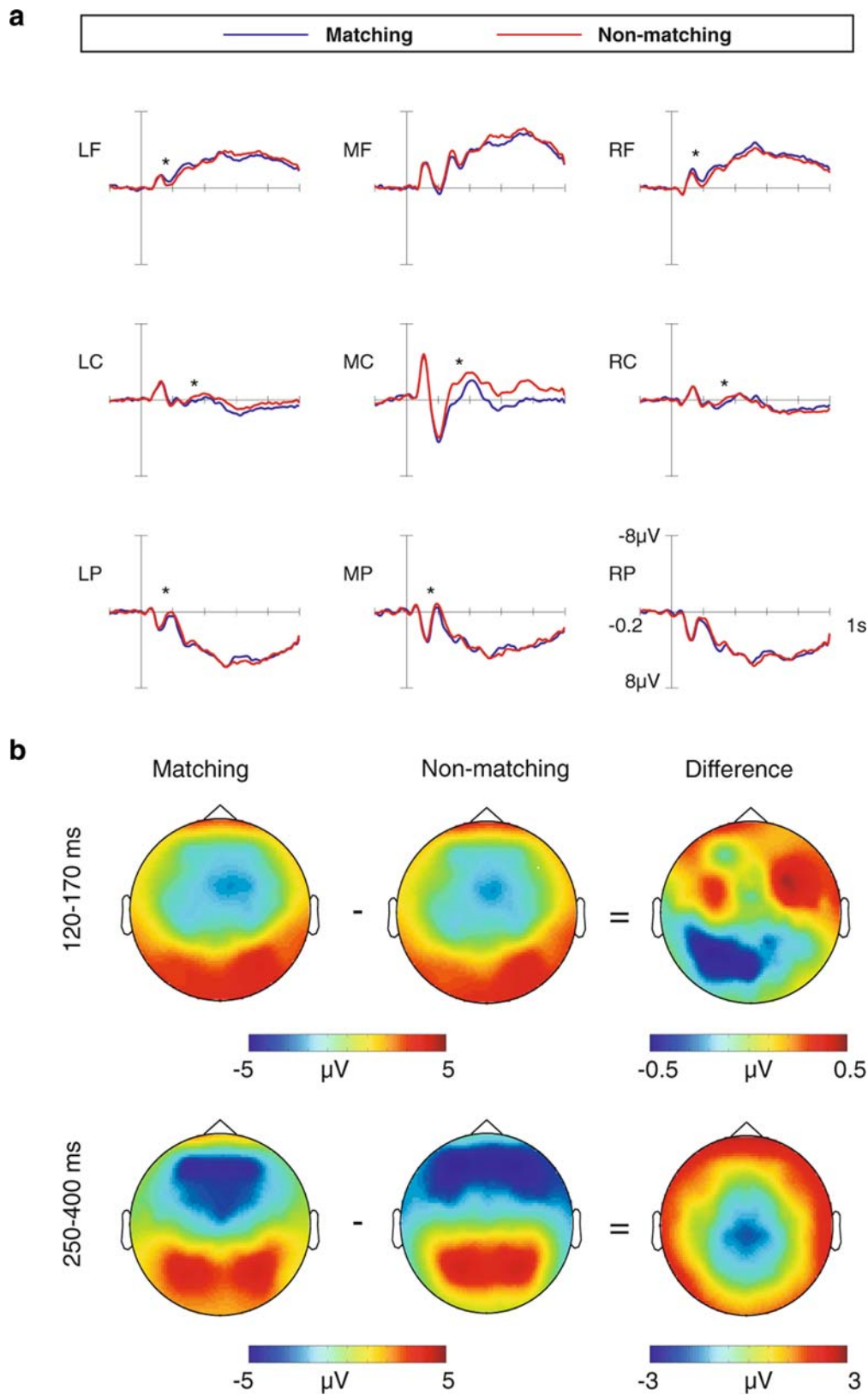


Fig. 3 a ERPs to auditory S2 inputs of semantically matching (blue graphs) and non-matching (red graphs) stimuli. Short latency effects were found at the left- and right-frontal and left- and middle-posterior ROI starting after around 120 ms. In addition, significant differences at

longer latencies were observed at medial-central as well as left- and right-central ROIs. **b** Topography of the short latency (upper panel) and long latency (lower panel) effects of semantic matching. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

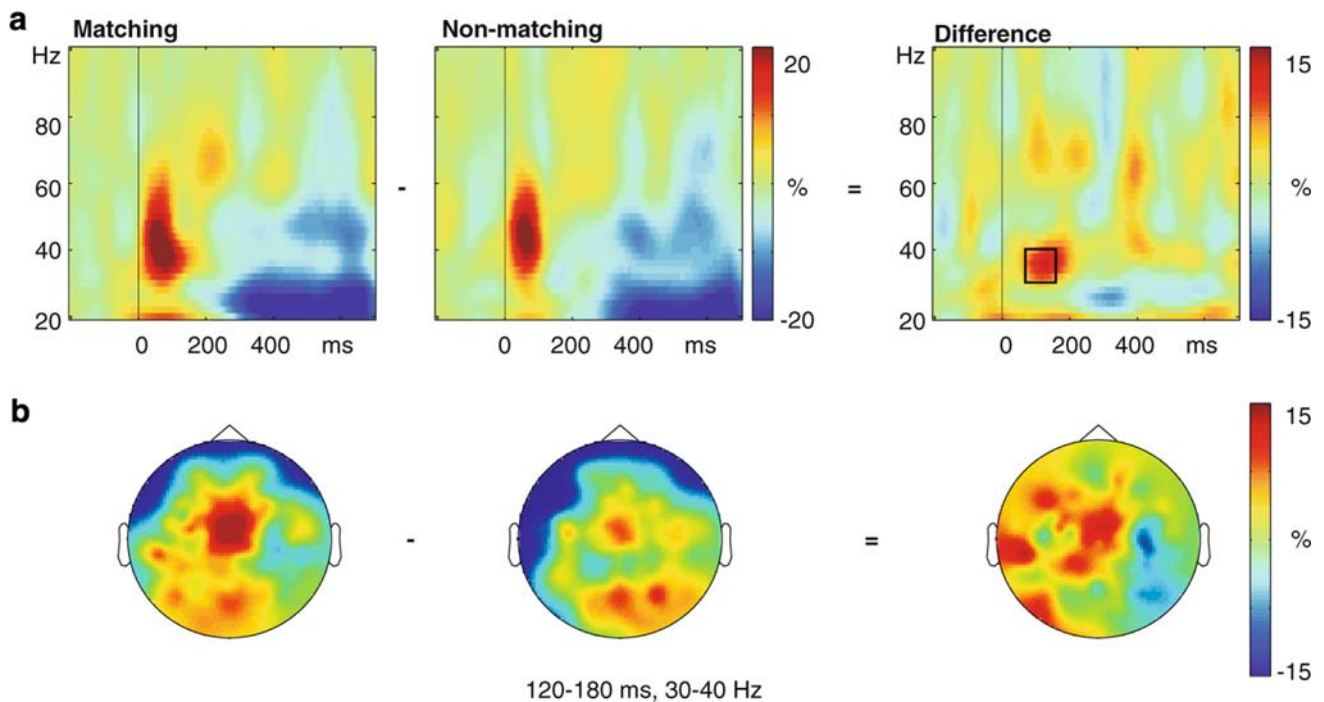


Fig. 4 a Time–frequency representations of total GBA in response to auditory (S2) stimuli at a medial-central region. An enhanced GBA to semantically matching compared to semantically non-matching stimuli

was observed after about 150 ms in the frequency range of 30–40 Hz. **b** Topography of total GBA (30–40 Hz, 120–180 ms) for semantically matching (*left map*) and non-matching (*right map*) auditory inputs

for the matching compared to non-matching condition. Follow-up ANOVAs revealed a larger GBA for matching compared to non-matching stimuli at the medial-central ROI (Fig. 3; $F_{1,18} = 13.78$, $p < 0.01$), and left-frontal ROI ($F_{1,18} = 10.16$, $p < 0.01$), suggesting that total GBA is linked to the matching of semantic inputs across modalities in WM. No significant effects were found for the other ROIs.

In line with previous reports (Senkowski et al. 2005, 2007b) an early enhancement of evoked GBA with a medial-central topography was observed for the responses to auditory S2 stimuli in both conditions (Fig. 5). The ANOVA for the 50–100 ms interval, however, did not reveal significant differences for the evoked GBA in response to matching and non-matching trials ($F < 1$). To further examine whether the above-described effect in the 120–180 ms time interval was specific for total GBA, an ANOVA was computed using the factors Activity Type (evoked, total power), Matching (matching, non-matching) and ROI (9 regions). This ANOVA revealed a three-way interaction between Activity Type, Matching, and ROI ($F_{8,144} = 2.41$, $p = 0.02$). Follow-up ANOVAs, which were computed for each of the 9 ROIs separately, showed an interaction between Activity Type and Matching at the left frontal region ($F_{1,18} = 9.202$, $p < 0.01$) and at the middle central region ($F_{1,18} = 11.698$, $p < 0.01$). This demonstrates that the effect of semantic matching is in primarily expressed in modulations of the total GBA.

Discussion

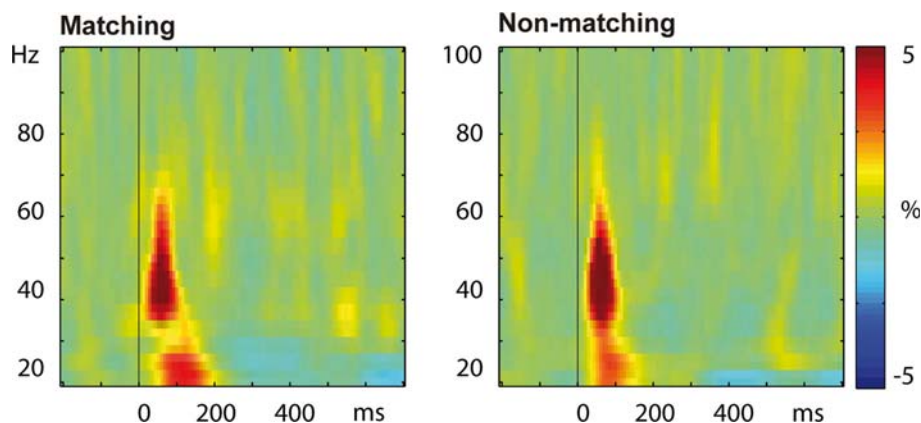
Here we examined the effects of semantic congruency on WM processing in a delayed visual-to-auditory matching paradigm. The central findings were effects of semantic congruency on GBA and ERPs that started around 120 ms after the auditory onset. The results will be discussed in detail below.

Memory matching in behavioral data

An interesting observation is the absence of effects of semantic matching on RTs. Previous implicit crossmodal priming studies have shown shorter RTs for semantically matching compared to semantically non-matching stimuli (Holcomb and Anderson 1993; Holcomb et al. 2005; Schneider et al. 2008b). Of particular note are the shorter RTs for semantically matching compared to semantically non-matching trials in a recent *implicit* visual-to-auditory priming study (Schneider et al. 2008b) that used the same experimental setup as used in the present experiment. This suggests that the absence of effects of semantic matching on RTs in the present study is related to the *explicit* matching of visual and auditory information in WM.

Another interesting finding is the enhanced error rate for semantically matching compared to semantically non-matching trials (i.e., 22 vs. 6%, respectively). The higher

Fig. 5 Time–frequency representations of evoked GBA in response to auditory (S2) stimuli at a medial-central region. An enhancement of evoked GBA was found at a latency of about 70 ms. There were no significant differences in evoked GBA between matching (*left map*) and non-matching (*right map*) trials



error rate for semantically matching trials may be due to the large ambiguity of the environmental sounds (e.g., Fabiani et al. 1996). It can be assumed that participants in the present study were able to identify the visual inputs without much effort, whereas the processing of the object sounds was presumably more difficult (Schneider et al. 2008a). Moreover, for semantically matching trials, there was only one single sound that matched the visual object, while there were a large number of possible visual-auditory object combinations for the non-matching trials. The observation of higher error rates for semantically matching trials is also in agreement with findings from an implicit visual-to-auditory priming study, in which environmental sounds and words were used as stimulus material (Orgs et al. 2006).

Since the error rates were higher in semantically matching compared to semantically non-matching stimuli, the lack of effects of semantic matching on RTs may be related to the higher difficulty for the processing of matching inputs. Indeed, when statistically controlling for the influence of the differences in error rates between matching and non-matching trials on RTs in an ANCOVA, an effect of semantic matching, similar to those reported in implicit multisensory priming studies (Holcomb and Anderson 1993; Holcomb et al. 2005; Schneider et al. 2008b) was found. Future studies will have to explore in more detail how differences in experimental tasks, i.e., *implicit* or *explicit*, influence behavioral performance in visual-to-auditory object matching paradigms.

Memory matching reflected by event-related potentials

The earliest modulations by semantic matching in the ERP occur over frontal and posterior scalp, at a latency of 120–170 ms after sound onset. The frontal negative deflection and the posterior positive deflection are larger for matching compared to non-matching stimuli. Although the topographic distribution of the ERP does not necessarily corre-

spond to the location of the underlying neural sources, some conclusions about the cortical structures linked to the reported effects may be drawn. The effects at left-frontal and right-frontal scalp possibly reflect an involvement of frontal and/or prefrontal brain regions during the semantic matching in WM. Frontal and prefrontal brain regions are linked to contextual matching between simultaneously presented semantic auditory and visual stimuli (Laurienti et al. 2003). The widespread effects at posterior scalp, on the other hand, may be indicative for processes in more posterior multisensory structures like superior temporal sulcus and/or lateral-occipital complex (Beauchamp et al. 2004; Molholm et al. 2004). Overall, the topography of the early ERP modulations by semantic matching suggests an involvement of higher cortical areas. Interestingly, no early effects in the ERP were observed when the same experimental setup is used in an *implicit* visual-to-auditory priming paradigm, in which the visual inputs were not relevant for the processing of auditory stimuli (Schneider et al. 2008b). This suggests a specific link of these early ERP effects to the *explicit* matching of meaningful visual and auditory information in WM.

The longer latency effect of semantic matching in the 250–400 ms interval is reflected by an ongoing negative deflection with a medial-central to medial-frontal topography. This deflection, which is frequently labeled as N400 component (Kutas and Hillyard 1980b), is stronger for semantically non-matching than for matching trials. Interestingly, our effect closely resembles an effect of semantic congruency that was recently observed in an implicit visual-to-auditory priming paradigm (Schneider et al. 2008b). We therefore propose that the longer latency effect may reflect a task-independent contextual integration process between the visual S1 inputs and the auditory S2 inputs. In line with previous findings for the semantic matching of language stimuli (Kutas and Federmeier 2000), the N400 component in the present study is larger for the processing of non-matching compared to the processing of

matching inputs. The observation of N400 modulations during the matching of auditory and visual inputs further supports the notion that this component is associated with the neural integration of contextual information across modalities (Orgs et al. 2006; Schneider et al. 2008b).

Memory matching reflected by gamma-band activity

The finding of enhanced GBA for semantic matching compared to non-matching inputs fits well with the hypothesis that activity in the gamma-band is important for memory matching processes (Gruber et al. 2004; Herrmann et al. 2004a; Jensen et al. 2007). For instance, the presentation of familiar environmental sounds with a long-term memory representation leads to higher total GBA compared to the presentation of non-familiar sounds without a long-term memory representation (Lenz et al. 2007). Another study demonstrates that auditory words which are presented in a study phase evoke higher total GBA in a subsequent test phase than newly presented words (Gruber et al. 2004). This suggests that the matching of auditory inputs with information stored in memory is paralleled by an increase in GBA. Similarly, in the present study semantically matching auditory inputs evoke higher total GBA than semantically non-matching stimuli. As such, our finding fits well with the match-and-utilization model (Herrmann et al. 2004b), which predicts enhanced GBA when bottom-up stimuli matches the template held in WM. Notably, our results also suggest that this model is applicable for neural processing across sensory modalities.

Of particular note is the finding that the effects on GBA occurred in the total but not in the short latency evoked responses. Recent studies have shown that multisensory integrative processing is reflected by an enhancement of evoked GBA (Senkowski et al. 2005, 2007b; Widmann et al. 2007). For instance, in a symbol-to-sound matching paradigm an enhanced *evoked* GBA was observed for auditory inputs that matched the elements of a visual pattern compared to auditory stimuli that did not match the visual pattern (Widmann et al. 2007). Similar to the present study, however, an enhancement of *total* GBA was observed around 100–200 ms after auditory onset. We suggest that the absence of effects of semantic matching on the short latency evoked GBA in our study might be related to the complex stimulus features of the environmental sounds. Studies that have reported multisensory effects on the evoked GBA have mostly used basic auditory stimuli like sinusoidal tones (Senkowski et al. 2005, 2007b) or triangle waves (Widmann et al. 2007). In contrast, when complex visual objects and auditory sounds were presented, multisensory effects were frequently observed for longer latency total GBA (Kaiser et al. 2005; Schneider et al. 2008a).

Conclusion

The present study demonstrates that multisensory visual-to-auditory object matching in WM is reflected by modulations of total GBA and ERPs, starting around 120 ms after the onset of semantically meaningful auditory stimuli. This suggests that synchronization of neural activity in the gamma-band expresses the degree of semantic matching in WM. Future studies may investigate whether the effect of semantic matching also occurs when auditory inputs are presented as S1 and visual as S2. Our observation that multisensory WM matching is reflected by total GBA is also in agreement with the hypothesis that binding by neural coherence is crucial for the integrative processing of sensory inputs across modalities (Kayser et al. 2008; Maier et al. 2008; Senkowski et al. 2008). Although our results do not allow for conclusions as to whether there may be modulations in neural coherence across cortical structures, the present findings suggest that there are local changes in gamma-band power with relation to the matching of semantic information across modalities.

Acknowledgments We thank Kriemhild Saha and Inga Fitzner for help during data acquisition and Inga Schepers and Hanna Krause for help during data acquisition and analysis. D.S. received support from a NARSAD 2006 young investigator award and the German Research Foundation (SE 1859/1-1). A.K.E. acknowledges support by the European Union (IST-2005-027268, NEST-PATH-043457, HEALTH-F2-2008-200728), the German Research Foundation (GRK 1247/1), the German Federal Ministry of Education and Research (01GW0561) and the Volkswagen Foundation (II/80609).

References

- Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41:809–823
- Debener S, Makeig S, Delorme A, Engel AK (2005) What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Brain Res Cogn Brain Res* 22:309–321
- Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 134:9–21
- Elliott R, Dolan RJ (1999) Differential neural responses during performance of matching and nonmatching to sample tasks at two delay intervals. *J Neurosci* 19:5066–5073
- Engel AK, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci* 2:704–716
- Fabiani M, Kazmerski VA, Cycowicz YM, Friedman D (1996) Naming norms for brief environmental sounds: effects of age and dementia. *Psychophysiology* 33:462–475
- Ferree TC, Luu P, Russel GS, Tucker DM (2001) Scalp electrode impedance, infection risk, and EEG data quality. *Clin Neurophysiol* 112:536–544
- Gruber T, Tsivilis D, Montaldi D, Muller MM (2004) Induced gamma band responses: an early marker of memory encoding and retrieval. *Neuroreport* 15:1837–1841

- Guthrie D, Buchwald JS (1991) Significance testing of difference potentials. *Psychophysiology* 28:240–244
- Herrmann CS, Lenz D, Junge S, Busch NA, Maess B (2004a) Memory-matches evoke human gamma-responses. *BMC Neurosci* 5:13
- Herrmann CS, Munk MH, Engel AK (2004b) Cognitive functions of gamma-band activity: memory match and utilization. *Trends Cogn Sci* 8:347–355
- Holcomb PJ, Anderson JE (1993) Cross-modal semantic priming: a time-course analysis using event-related potentials. *Lang Cogn Process* 8:33
- Holcomb PJ, Anderson J, Grainger J (2005) An electrophysiological study of cross-modal repetition priming. *Psychophysiology* 42:493–507
- Jensen O, Kaiser J, Lachaux JP (2007) Human gamma-frequency oscillations associated with attention and memory. *Trends Neurosci* 30:317–324
- Jokisch D, Jensen O (2007) Modulation of gamma and alpha activity during a working memory task engaging the dorsal or ventral stream. *J Neurosci* 27:3244–3251
- Kaiser J, Hertrich I, Ackermann H, Mathiak K, Lutzenberger W (2005) Hearing lips: gamma-band activity during audiovisual speech perception. *Cereb Cortex* 15:646–653
- Kayser C, Petkov C, Logothetis NK (2008) Visual modulation of neurons in auditory cortex. *Cereb Cortex* 18:1560–1574
- Kutas M, Federmeier KD (2000) Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn Sci* 4:463–470
- Kutas M, Hillyard SA (1980a) Reading between the lines: event-related brain potentials during natural sentence processing. *Brain Lang* 11:354–373
- Kutas M, Hillyard SA (1980b) Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207:203–205
- Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53:279–292
- Laurienti PJ, Wallace MT, Maldjian JA, Susi CM, Stein BE, Burdette JH (2003) Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. *Hum Brain Mapp* 19:213–223
- Lenz D, Schadow J, Thaeerig 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* 64:31–38
- Lutzenberger W, Ripper B, Busse L, Birbaumer N, Kaiser J (2002) Dynamics of gamma-band activity during an audiospatial working memory task in humans. *J Neurosci* 22:5630–5638
- Maier JX, Chandrasekaran C, Ghazanfar AA (2008) Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Curr Biol* 18:963–968
- 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
- Orgs G, Lange K, Dombrowski JH, Heil M (2006) Conceptual priming for environmental sounds and words: an ERP study. *Brain Cogn* 62:267–272
- Schneider T, Engel AK, Debener S (2008a) Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Exp Psychol* 55:121–132
- Schneider TR, Debener S, Oostenveld R, Engel AK (2008b) Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *Neuroimage* 42:1244–1254
- Senkowski D, Talsma D, Herrmann CS, Woldorff MG (2005) Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. *Exp Brain Res* 3–4:411–426
- Senkowski D, Molholm S, Gomez-Ramirez M, Foxe JJ (2006) Oscillatory beta activity predicts response speed during a multisensory audiovisual reaction time task: a high-density electrical mapping study. *Cereb Cortex* 16:1556–1565
- Senkowski D, Saint-Amour D, Kelly SP, Foxe JJ (2007a) Multisensory processing of naturalistic objects in motion: a high-density electrical mapping and source estimation study. *Neuroimage* 36:877–888
- Senkowski D, Talsma D, Grigutsch M, Herrmann CS, Woldorff MG (2007b) Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia* 45:561–571
- Senkowski D, Schneider TR, Foxe JJ, Engel AK (2008) Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci* 31:401–409
- Siegel M, Donner TH, Oostenveld R, Fries P, Engel AK (2007) High-frequency activity in human visual cortex is modulated by visual motion strength. *Cereb Cortex* 17:732–741
- Stekelenburg JJ, Vroomen J (2007) Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cogn Neurosci* 19:1964–1973
- 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
- Widmann A, Gruber T, Kujala T, Tervaniemi M, Schroger E (2007) Binding symbols and sounds: evidence from event-related oscillatory gamma-band activity. *Cereb Cortex* 17:2696
- Womelsdorf T, Schoffelen JM, Oostenveld R, Singer W, Desimone R, Engel AK, Fries P (2007) Modulation of neuronal interactions through neuronal synchronization. *Science* 316:1609–1612