

Intermodal attention affects the processing of the temporal alignment of audiovisual stimuli

Durk Talsma · Daniel Senkowski · Marty G. Woldorff

Received: 31 October 2008 / Accepted: 12 May 2009 / Published online: 4 June 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract The temporal asynchrony between inputs to different sensory modalities has been shown to be a critical factor influencing the interaction between such inputs. We used scalp-recorded event-related potentials (ERPs) to investigate the effects of attention on the processing of audiovisual multisensory stimuli as the temporal asynchrony between the auditory and visual inputs varied across the audiovisual integration window (i.e., up to 125 ms). Randomized streams of unisensory auditory stimuli, unisensory visual stimuli, and audiovisual stimuli (consisting of the temporally proximal presentation of the visual and auditory stimulus components) were presented centrally while participants attended to either the auditory or the visual modality to detect occasional target stimuli in that modality. ERPs elicited by each of the contributing sensory modalities were extracted by signal processing techniques from the combined ERP waveforms elicited by the multisensory stimuli. This was done for each of the five different

50-ms subranges of stimulus onset asynchrony (SOA: e.g., V precedes A by 125–75 ms, by 75–25 ms, etc.). The extracted ERPs for the visual inputs of the multisensory stimuli were compared among each other and with the ERPs to the unisensory visual control stimuli, separately when attention was directed to the visual or to the auditory modality. The results showed that the attention effects on the right-hemisphere visual P1 was largest when auditory and visual stimuli were temporally aligned. In contrast, the N1 attention effect was smallest at this latency, suggesting that attention may play a role in the processing of the relative temporal alignment of the constituent parts of multisensory stimuli. At longer latencies an occipital selection negativity for the attended versus unattended visual stimuli was also observed, but this effect did not vary as a function of SOA, suggesting that by that latency a stable representation of the auditory and visual stimulus components has been established.

Keywords Electrophysiology · EEG · ERP · Multisensory · SOA

D. Talsma
Cognitive Psychology Department, Vrije University Amsterdam,
Amsterdam, The Netherlands

D. Talsma (✉)
Department of Cognitive Psychology and Ergonomics,
University of Twente, PO Box 215,
7500 AE Enschede, The Netherlands
e-mail: d.talsma@utwente.nl

D. Senkowski
Department of Neurophysiology and Pathophysiology,
University Medical Center Hamburg-Eppendorf,
Hamburg, Germany

D. Talsma · M. G. Woldorff
Center for Cognitive Neuroscience and Department of Psychiatry,
Duke University, Durham, NC, USA

Introduction

To understand the processing of audiovisual stimuli, it is useful to study the conditions in which the audiovisual inputs break apart into separate auditory and visual perceptual processes. For example, in a badly mastered audio track of a movie, there can be a noticeable desynchronization between visual and auditory information streams. Such a desynchronization can also be observed in real life, such as in the case of a distant thunderstorm, a music concert in a large arena, or a fast jet aircraft that appears to fly ahead of its sound. These latter desynchronization phenomena occur

because light travels at substantially faster speed than sound, thus causing the visual inputs to reach the visual receptors considerably earlier than the auditory inputs reach the auditory ones.

Behavioral studies have shown that auditory and visual stimuli coming from the same location were judged most likely to originate from a common cause when visual stimuli precede the auditory by about 50 ms (even as compared to simultaneous occurrence) (Lewald et al. 2001; Lewald and Guski 2003). Presumably, this optimum timing difference reflects the fact that the brain has evolved to process multisensory stimuli that are located some distance away from the observer. Due to the difference between the velocities of sound and light, auditory signals from an object located at a distance of about 15 m (probably representing a reasonable average distance of real-world objects), reach our ears approximately 50 ms after light reaches the retina. To process such objects optimally, our brain has presumably evolved to compensate for this difference by means of a higher neural transmission rate for auditory signals than for visual ones. Moreover, it has been shown that the temporal window for multisensory integration processes can be adapted on the basis of viewing distance (King 2005; Spence and Squire 2003; but see Lewald and Guski 2004) or exposure (Navarra et al. 2005), and that multisensory perception appears to stay relatively intact even with fairly large temporal discrepancies between visual and auditory stimuli (Shams et al. 2001; Shams et al. 2002). These and other results (Fendrich and Corballis 2001; Morein-Zamir et al. 2003; Vroomen and De Gelder 2004; Talsma et al. 2007; Busse et al. 2005) have led to the suggestion that the processing of visual and auditory inputs may be influenced by their relative timing, as well as by endogenous factors, such as attention.

Although stimulus processing effects as a function of temporal alignment are not well understood, some evidence exists which indicates that visual inputs are perceptually realigned on the basis of an auditory stimulus (DiLuca et al. 2009). For example, when the temporal characteristics of auditory and visual stimuli do not match, perceptual reports regarding the temporal characteristics of visual stimuli tend to be biased toward the temporal characteristics of the auditory stream (Shams et al. 2002). More specifically, when participants were presented two flashes accompanied by three tone pips, they tended to report seeing three flashes rather than just two. We (Senkowski et al. 2007a) and others (e.g., Andersen and Mamassian 2008; Spence and Squire 2003) have previously reported that participants begin to detect onset differences between auditory and visual stimuli when they are at least 125–150 ms apart, and only able to accurately report these onset differences when the visual and auditory stimuli are 250 ms or more apart. These findings suggest that there is a relatively broad integration time

window of as large as 250 ms, in which stimuli from different modalities typically tend to be integrated into a single multisensory percept. In addition, animal physiology research has revealed that a large number of neurons in the superior colliculus show enhanced responsiveness when visual and auditory stimuli are presented within a time window of less than 100–200 ms (Meredith et al. 1987; Meredith 2002).

The latter finding is a special case of a more general result from the animal physiological literature: a large number of neurons in superior colliculus elicit much larger responses to multisensory (i.e., audiovisual: AV) stimuli than they would do in response to the unisensory auditory (A) or unisensory visual (V) component parts presented separately (Stein and Meredith 1993; Wallace et al. 1993). Supplementing these findings, human electrophysiology studies have revealed multisensory interactions in the early event-related potentials (ERPs) over sensory as well as non-sensory areas of the cortex, although how early this can occur and under what circumstances has varied (Giard and Peronn t 1999; Molholm et al. 2002; Teder-S lej rvi et al. 2002; Talsma et al. 2007; Senkowski et al. 2007a).

Nevertheless, it still remains a question whether and in what way multisensory processing alters as a function of the relative timing between visual and auditory inputs, even when both visual and auditory inputs of the multisensory stimulus fall within the integration window. As noted above, behavioral and animal physiological studies have found that the window of integration is relatively broad. Yet at the same time behavioral studies have shown that there appears to be an optimum relative stimulus timing wherein visual stimuli precede auditory stimuli by about 50–100 ms, such that these stimuli are subjectively most likely as being perceived as simultaneous (Lewald and Guski 2003). In addition, other behavioral studies have shown that, when participants are required to judge the temporal order of auditory and visual stimuli, they are less able to make an accurate judgment when these stimuli were presented at or near this optimum point of subjective simultaneity (i.e., when the visual stimulus preceded the auditory stimulus by about 50–100 ms; Keetels and Vroomen 2005; Vroomen and Keetels 2006; McDonald et al. 2005). Interestingly, these temporal order judgments gradually shift toward the direction of the actual temporal order as the auditory and visual stimulus inputs deviate more and more from the point of subjective simultaneity. This suggests that neural processes underlying these audiovisual interactions would also gradually alter with increasing temporal offsets between auditory and visual stimuli, despite still falling within the integration window.

The primary goal of the present study was to investigate to what degree these processes involved in resolving temporal asynchrony are affected by intermodal attention. Previous studies have shown that attention can interact with

processes involved in multisensory perception (Talsma and Woldorff 2005a; Senkowski et al. 2005). However, these effects have generally been investigated under conditions in which stimuli were fully attended or unattended; that is, attention was directed to the multisensory stimuli (i.e., at both visual and auditory modalities simultaneously), or directed elsewhere. A more recent study suggests a more complex relationship between attention and audiovisual processing (Talsma et al. 2007). Whereas we found that it is indeed a requirement of attention to be focused on both modalities simultaneously for the relatively early latency (~50 ms after stimulus onset) multisensory interactions to occur, longer-latency effects suggested the spreading of attention from the visual modality to the auditory under conditions in which only the visual modality was attended (Busse et al. 2005). Specifically, when the visual modality was attended, a slow, long-latency (>300 ms) fronto-central negativity was observed in response to multisensory stimuli, but not to either the corresponding unisensory visual or unisensory auditory stimuli occurring alone. This fronto-central negativity strongly resembled the intramodal auditory attention effect known as the late processing negativity, which is elicited by attended auditory stimuli, and which is believed to reflect the prolonged processing of relevant auditory stimuli (Näätänen 1992).

ERPs are ideally suited to track the temporal characteristics of psychologically related physiological processes, enabling them to be very effective in assessing how and when interactions between closely timed auditory and visual stimuli occur. Thus far, however, only a few electrophysiological studies have systematically addressed the effects of temporal asynchrony between visual and auditory inputs on the electrophysiology of multisensory processing (e.g., Senkowski et al. 2007b; Stekelenburg and Vroomen 2005).

One reason why only few studies have investigated such timing effects is likely related to the problem induced by the overlap between the ERPs elicited by the visual and the auditory stimulus inputs that occur closely in time. More specifically, since both the visual and auditory inputs of a multisensory audiovisual event will each evoke an ERP, the time-locked averages to the stimuli in one modality will be contaminated by an overlapping ERP elicited by the other-modality stimulus. If the goal is to study the effects on stimulus processing of varying the relative timing by auditory and visual inputs, then the latency range and structure of this overlap distortion will differ systematically across the timing asynchrony conditions, thus confounding assessment of the actual interaction effects on multisensory processing (Woldorff 1993).

Stekelenburg and Vroomen (2005) addressed the overlap problem by including the corresponding unisensory auditory stimuli in the sequence and using the ERPs as estimates for the overlapping activities. They reported a

decrease in the posterior N1 when the auditory component preceded the visual one, compared to when visual stimuli were presented alone, and a similar, albeit smaller amplitude reduction of the N1 when auditory and visual inputs were presented simultaneously. The authors reported these results relative to the onset of the visual input using three fixed temporal asynchronies between visual and auditory inputs: -100 ms (auditory preceding visual), 0 ms (simultaneous presentation of visual and auditory inputs), and +100 ms (auditory inputs lagging behind the visual).

We have previously observed multisensory audiovisual interactions in the evoked high-frequency oscillatory gamma-band responses (GBRs, 30–80 Hz) (Senkowski et al. 2007b) across this range using a fine range of temporal asynchronies (see also Fig. 1b). In the present paper we use a decomposition analysis based on the Adjar filtering technique (Woldorff 1993)—a deconvolution version of the simple subtraction method of Stekelenburg and Vroomen (2005)—that takes ERPs elicited by unisensory stimuli as estimates for the overlapping ERPs (see “Methods” for details). This analysis was applied to the ERP responses from the same dataset for which Senkowski et al. (2007b) analyzed effects on the high-frequency evoked GBRs. More specifically, visual and auditory inputs of multisensory stimuli were presented in close succession (relative asynchronies ranging between -125 ms and 125 ms). Participants were instructed to attend to only one modality during the length of a trial block and to detect infrequent target stimuli in the designated modality. These target stimuli were slightly different from the regular non-target stimuli. In this paper, we specifically focus on reporting ERP data that were evoked by the visual stimuli as a function of attention and the audiovisual synchrony, for which we have found the most pronounced effects.

Since all stimuli were presented from central locations, we expected to extend some previous findings of non-spatial intermodal attention (De Ruyter et al. 1998; Talsma and Kok 2001; Talsma et al. 2007). Most notably, we expected that intermodal attention during multisensory stimulation would affect the early P1 and N1 visual sensory components (cf. Talsma et al. 2007), being larger when the visual modality versus the auditory modality was attended. In addition, we expected that intermodal attention to visual stimuli would be reflected in an endogenous ERP component, known as the occipital selection negativity, which is a component that has been reported to reflect intramodal (e.g., Kenemans et al. 1993), as well as intermodal attentional selection processes (Talsma and Kok 2001). Particular interest was on how these various intermodal attention effects on the processing of a visual stimulus component would vary if that component occurred as part of an audiovisual multisensory stimulus with differing degrees of asynchrony of the audio and visual parts.

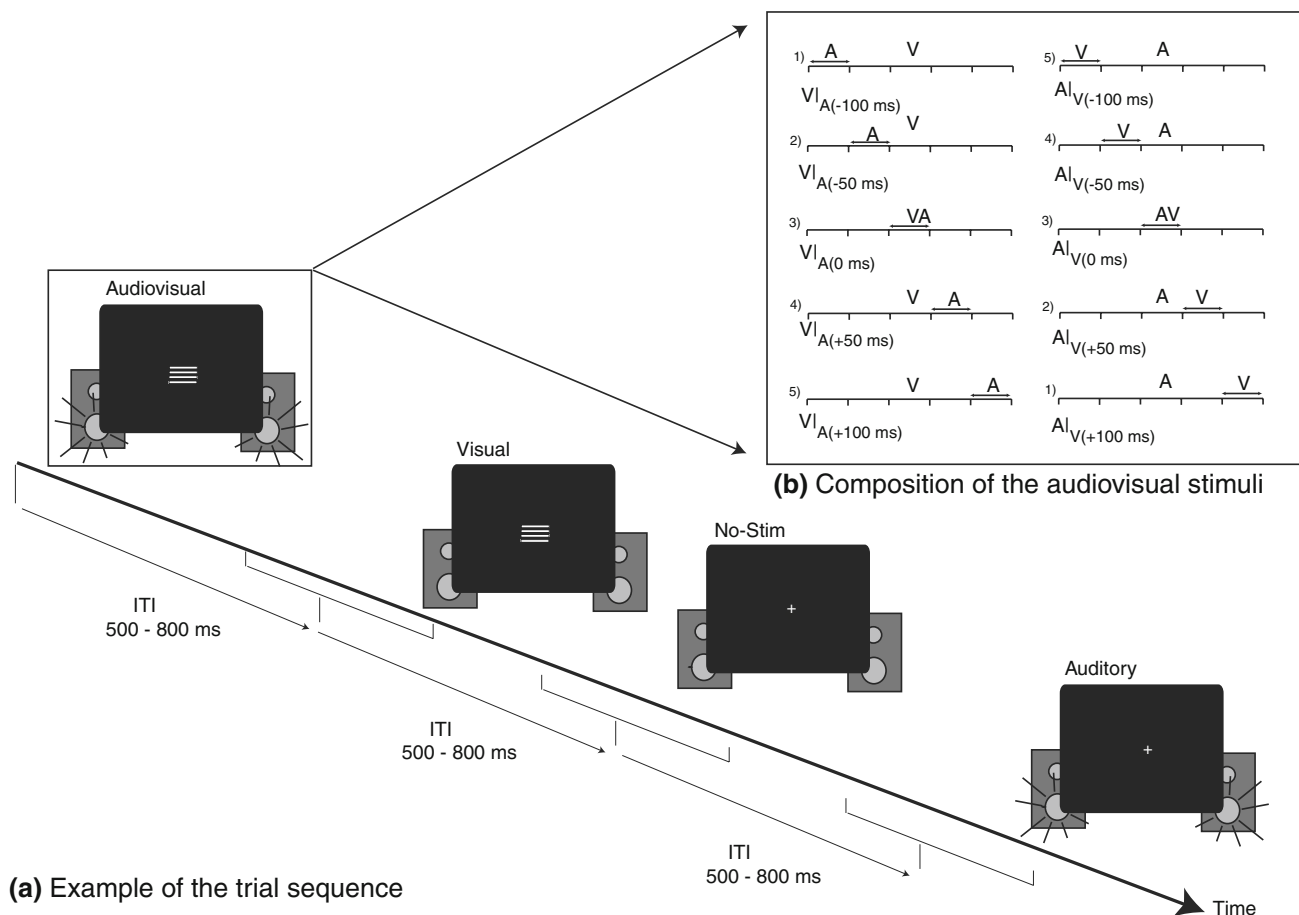


Fig. 1 General layout of the paradigm. **a** Example of a short excerpt from a possible trial sequence. Visual, auditory, multisensory, or no-stim trials were presented with random ITIs of 500–800 ms. The presentation order of the various stimulus types was fully randomized. **b** Decomposition of the relative timing of the auditory and visual stimuli within each of the subtypes of audiovisual stimuli, when viewed as

time-locked to either the visual input (left) or the auditory input (right). Notice that there were actually five different multisensory stimuli. For instance, $V|_{A(-100\text{ ms})}$ trials were the same as the $A|_{V(+100\text{ ms})}$ trials, but viewed as time-locked to either the visual input (left panel) or the auditory input (right panel)

Methods

Participants

Twenty healthy participants volunteered for the experiment (age: 25.2; 9 males). All participants were recruited through advertisements at the campus of Duke University and were paid \$10 per hour. All participants were right handed and had normal or corrected-to-normal vision and gave written informed consent to volunteer for the study. No one reported having a history of neurological or psychiatric disorder.

Task and stimuli

Visual stimuli consisted of white horizontal gratings, presented on a black background (5.5×4.9 cm, subtending a visual angle of $5.4 \times 4.8^\circ$) for a duration of 100 ms. Auditory stimuli consisted of 1600 Hz sinusoidal tone pips, pre-

sented at a sound pressure level of 65 dB(A) for a duration of 100 ms (including linear rise and fall times of 10 ms). These auditory stimuli were presented using two speakers that were placed behind the computer screen used for visual presentation, positioned such that the subjective location of the auditory stimuli matched that of the visual stimuli. Audiovisual stimulus trials were presented with stimulus onset asynchronies (SOAs) of the auditory and visual inputs ranging randomly from trial to trial between ± 125 ms (see Fig. 1b). Importantly, for the analyses we divided the ± 125 ms SOA range into five different 50-ms subranges. The audiovisual SOAs were randomized at 1 ms resolution, but restricted so that an equal number of trials were presented in each 50-ms SOA condition. In addition to the above-described stimuli, which were designated as the standard (i.e., non-target) stimuli, a small proportion of the stimuli were characterized by an additional stimulus feature, which enabled them to serve as target stimuli in the stimulus streams. These deviant stimuli were highly similar

to the standards but contained a transient dip in brightness (visual) or volume (auditory) halfway through the duration of the stimulus, inducing the subjective impression of the stimulus appearing to flicker (visual target) or to stutter (auditory target). The degree of the intensity reduction was determined for each participant individually during a training session prior to the experiment (Senkowski et al. 2007b; Talsma and Woldorff 2005a), so that each deviant was just detectable for each participant. Multisensory deviants were composed of either a visual standard paired with an auditory deviant, or vice versa.

Stimuli were presented as a continuous stream of randomly mixed unisensory auditory, unisensory visual, and multisensory audiovisual stimulus events. Of the multisensory stimuli, a total number of 160 standard stimuli (consisting of a visual standard paired with an auditory standard) and 96 deviant stimuli (half consisting of the visual standard paired with the auditory deviant and the other half consisting of an auditory standard with a visual deviant) were presented with each of the five audio-visual SOA temporal-synchrony subranges. Likewise, 160 unisensory auditory standards, 160 unisensory visual standards, 96 unisensory auditory deviants, and 96 unisensory visual deviants were presented. In addition, 400 “no-stim” trials were included in each condition. These trials were used to estimate the between-trial ERP overlap (Talsma and Woldorff 2005a, b; see Gondan and Röder 2006, for a similar suggestion). No-stims are trials on which no actual stimulus is presented, but which are presented with the same randomization characteristics as the regular trials (Burock et al. 1998). Because the randomization of the no-stims trials satisfy the criteria delineated by Busse and Woldorff (2003) to not elicit electrophysiological responses themselves, they can be used to pick up overlapping ERP activity resulting from either anticipatory processes or from adjacent-trial ERPs (see Talsma and Woldorff 2005a, b for a full discussion) All stimulus trials were presented using relatively short inter-trial-intervals (ITIs) that ranged from 500 to 800 ms (mean 650 ms) and were equally divided across 16 blocks of trials.

Procedure

In each block, the participants’ task was to attend to a designated modality in each run and detect the infrequent deviant target stimuli in that modality. In one condition, they were required to attend to the auditory modality and respond to the detection of any auditory deviants, and in another condition they were required to attend only to the visual modality and respond only to the visual deviants. In both conditions, participants were instructed to report their detection of the target deviants by making a button-press response as fast as possible while maintaining high accu-

racy. The two attention conditions (attend-auditory, attend-visual) alternated every 2 blocks, with the starting condition being randomized across participants. One practice block for each attention condition was run prior to the experiment to familiarize participants with the procedure. Participants were further instructed to maintain fixation at a centrally presented cross during the run, as well as to minimize head and body movements.

Apparatus

The EEG and behavioral data were recorded in a dimly lit, sound attenuated, electrically shielded chamber. Stimulus presentation was controlled by a personal computer running the “Presentation” software package (Neurobehavioral Systems, Inc., Albany, CA, USA). EEGs were recorded from 64 equally spaced tin electrodes,¹ mounted in a customized elastic cap (Electro-Cap International, Inc) and referenced to the right mastoid during recording. Electrode impedances were kept below 2 k Ω for the mastoids and ground, 10 k Ω for the eye electrodes, and 5 k Ω for the remaining electrodes. Horizontal eye movements were monitored by two electrodes at the outer canthi of the eyes. Vertical eye movements and eye-blinks were detected by electrodes placed below the orbital ridge of both eyes, referenced to two electrodes directly located above the eyes. During recording, eye movements were also monitored using a closed circuit video monitoring system. EEG was recorded using a Neuroscan (SynAmps) acquisition system. All EEG channels were recorded with a band-pass filter of 0.01 to 100 Hz and a gain setting of 1,000. Raw signals were continuously digitized with a sampling rate of 500 Hz and digitally stored for off-line analysis.

Data analysis

Behavioral data

Reaction times (RTs) for correctly detected targets, target error rate, and false alarm rates were computed separately for the different conditions. RT and error rate measures

¹ The EEG was recorded by the following electrodes: FPz, Fz, FCz, Cz, Fp1 m, Fp2 m, F3a, F4a, F3 s, F4 s, FC1, FC2, C1a, C2a, F7a, F8a, F3i, F4i, C3a, C4a, PA1a, PA2a, F7p, F8p, C5a, C6a, T3', T4', LC, RC, LIO, RIO, Lm, Inz, Ozi, Ozs, Pzi, Pzs, I1, I2, O1i, O2i, O1', O2', PO1, PO2, P1', P2', C1p, C2p, T11, T12, TO1, TO2, P3i, P4i, P3a, P4a, C3', C4', T35i, T46i, C5p, and C6p. These electrode positions are named relative to their approximate 10–10 equivalents. A suffix of ‘s’ indicates that the electrode was placed slightly (i.e., within 1–1.5 cm) superior to the indicated standard position; ‘i’ indicates it was placed inferior to the standard position. Similarly, ‘a’ and ‘p’ indicate the electrode in question was positioned slightly anterior or posterior to the standard locations, respectively. Electrodes position <1 cm are named by the standard location name with an added prime mark (e.g., C4').

were submitted to multivariate analysis of variance (MANOVA) with SOA (5 levels: -100 , -50 , 0 , 50 , and 100 ms) and Attention (visual or auditory) as within-subject factors. In addition, the RTs and error rates of the multisensory stimuli were contrasted against the RTs and error rates of the unisensory stimuli, using pair-wise statistical comparisons.

ERP analyses

Artifact rejection was performed off-line by discarding epochs of the EEG that were contaminated by eye movements ($>\pm 130 \mu\text{V}/10$ ms) using the detection algorithms described in Talsma and Woldorff (2005b). Eye-blinks were corrected using a time-domain linear regression algorithm (e.g., Kenemans et al. 1991). Further artifact detection was conducted using an auto-adaptive averaging procedure that finds artifactual trials on the basis of signal to noise properties of the ERP waveform (Talsma 2008). Approximately one-third of the trials were discarded due to artifacts. The number of discarded trials did not differ significantly between attend-visual and attend-auditory blocks ($F < 1$). Prior to averaging, EEG data were band-pass

filtered between 0.03 and 25 Hz, using a half-Gaussian (causal) kernel. Averages were calculated for the different stimulus types from 1,000 ms before to 1,200 ms after stimulus onset. After averaging, all channels were re-referenced to the algebraic average of the two mastoid electrodes. For the multisensory stimuli, ERPs were computed relative to the onset of the visual input as well as to the onset of the auditory input.

Removal of overlapping adjacent stimulus activity

The ERP effects of multisensory processing were extracted using the following procedure (Fig. 2). The between-trial overlap, which resulted from the relatively fast trial presentation rate, was removed by subtracting the time-locked averages to the no-stim trials from the regular ERPs (Talsma and Woldorff 2005a, b). However, although this procedure corrects for the between-trial overlap, the ERPs for the audiovisual trials were still distorted by within-trial overlap because the visual and auditory components of these stimuli were presented in such close temporal proximity. The selective averaging time-locked to the inputs of one modality of these multisensory stimuli resulted in an

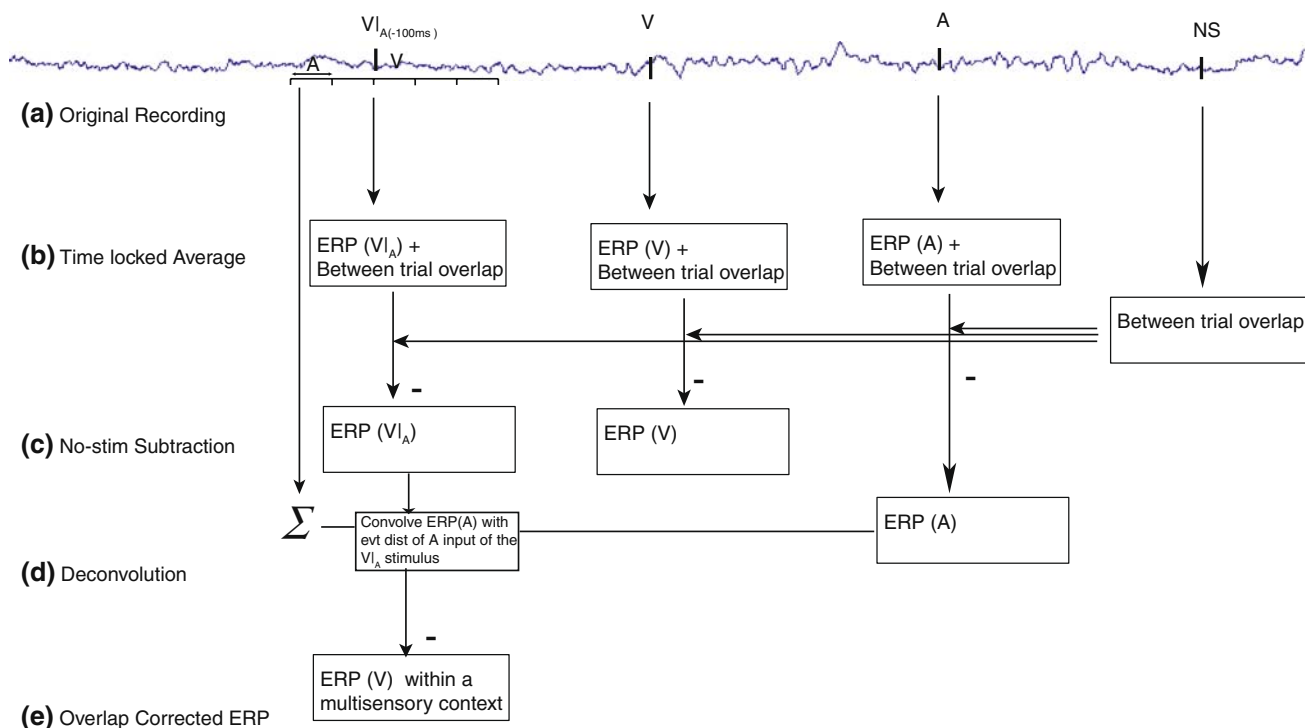


Fig. 2 Steps involved in extracting the ERP data. An example of removing the overlapping auditory component from a $V|_{A(-100\text{ms})}$ stimulus is shown here. A similar procedure was used for the other audiovisual stimulus types. After EEG recording (a), condition-wise averages (b) of all the multisensory ERPs (separately for each time window), unisensory stimuli, and no-stim (NS) trials were computed. c Subtraction of the no-stim ERPs from the unisensory and multisensory ERPs resulted in an ERP that eliminated the between-trial overlap,

but which was still composed of contributions from an auditory input and a visual input. d The ADJAR procedure (Woldorff 1993) was used to convolve the unisensory ERP (as an estimate for the overlapping activity) with the distribution of all immediately preceding events (evt dist) to obtain an overlap free estimate of the contribution of one single modality (e), plus some variations due to attention and/or multisensory interaction processes (e.g., “ERP(V) within a specific multisensory context”)

ERP waveform to that stimulus type, plus distortion due to the ERPs to the temporally proximal occurrence of the other-modality input. Because each time-locked average to the sensory input of one modality had the other-modality sensory input occurring across a 50-ms SOA subrange relative to the time-lock point (e.g., 25–75 ms after), the contribution of the overlapping ERP elicited by the other-modality input was partially “smeared out” and thus somewhat reduced in the time-locked average to the first modality. However, with such a relatively small temporal jitter, there would still be substantial overlap from the adjacent ERP activity (Woldorff 1993).

To remove this within-trial overlap in the multisensory ERPs, an adapted version of the adjacent response filter (ADJAR) procedure (Woldorff 1993) was used. ADJAR is a post-experimental correction technique to estimate and remove from the ERP waveform the overlap that results from adjacent trials in fast-rate sequences. This is accomplished by convolving an estimate of the overlapping waveforms with the event-distribution of the preceding and succeeding events. The procedure typically starts with averages that are somewhat distorted by overlap. After the first iteration of estimating and subtracting the distorting overlap, the ADJAR algorithm iteratively uses the partially corrected ERP waveforms for better overlap estimates until the solutions converge after multiple iterations. In the present case, however, the unisensory ERP averages were used as estimates for the overlapping ERP activity from the other-modality input. Specifically, this was done by convolving the unisensory other-modality ERP with the event-distribution of the overlapping other-modality stimulus input of the multisensory stimulus. As the unisensory ERP averages were not distorted by overlap here, only a single iteration of estimating and subtracting the adjacent-response overlap was necessary.

After overlap removal, the resulting ERP waveform consisted of the basic time-locked unisensory response, plus the interaction activity resulting from its occurrence within a multisensory (audiovisual) context. The extracted waveforms from the multisensory trials were very similar to the ERP traces of the corresponding *true* unisensory stimuli for that modality, but with some variations due to these multisensory interactions. Multisensory processing effects can therefore be investigated by comparing the extracted ERPs with the true unisensory ERP responses. These extracted ERPs are reported relative to the onset of the visual input (V) or relative to the auditory input (A) of a multisensory AV event, depending on the time-lock point. In the remainder of this paper, temporal relation of the other-modality stimulus onset time will be reported using a subscripted notation. For instance, a response relative to a visual stimulus, with an auditory stimulus *preceding* this visual stimulus in the 50-ms time window centered around 100 ms,

would be indicated as $V|_{A(-100\text{ ms})}$. Accordingly, an ERP time-locked to the visual stimulus with the auditory stimulus *succeeding* it with its 50-ms subrange centered around 100 ms would be notated as $V|_{A(100\text{ ms})}$.

Subsequent ERP analyses focused on the overlap-corrected ERPs to the standard stimuli. This included both the estimated single-modality ERPs extracted from the multisensory standard stimulus trials and the ERPs from the unisensory standard stimuli. Scalp topography maps and plots of the ERP traces were used for the analyses of multisensory interactions between auditory and visual ERP components. Based on previous reports, the statistical analysis of multisensory processing effects was focused on a selection of fronto-central areas (for effects that were mainly driven by the auditory modality), and one over posterior areas (for effects that were mainly driven by the visual stimuli).

As the effects of Attention and SOA and their interactions were substantially more robust on the extracted visual ERPs, this paper focuses on the components of these responses. Initial inspection of the visual ERPs extracted from the multisensory responses revealed the presence of several components that were modulated by intermodal attention. Of these components, we focus here on the posterior P1, N1, and P2 components of the extracted visual ERP responses at ~ 110 , ~ 150 , and ~ 250 after stimulus onset, respectively, as well as an anterior N1 component that occurs slightly later than the visual P1. The significance of these effects was assessed by computing mean amplitudes across a short window surrounding the peaks of these components, using electrodes that were located near the observed maxima of each component. These amplitude values were then submitted to a multivariate analysis of variance (MANOVA) that contrasted the amplitudes of these conditions in each multisensory condition against the amplitudes of the unisensory control.

In addition, effects of SOA and Attention within the multisensory stimuli were tested using a MANOVA containing the factors of SOA (5 levels, corresponding to the five different SOA ranges), and Attention (two levels: evoking input was attended or unattended). Specific details regarding latency and electrode selection is given in the respective section, where each component is described.

Results

Behavioral results

Reaction times

Figure 3a shows the mean response times (RTs) for the target stimuli for each of the SOA-subrange conditions, along

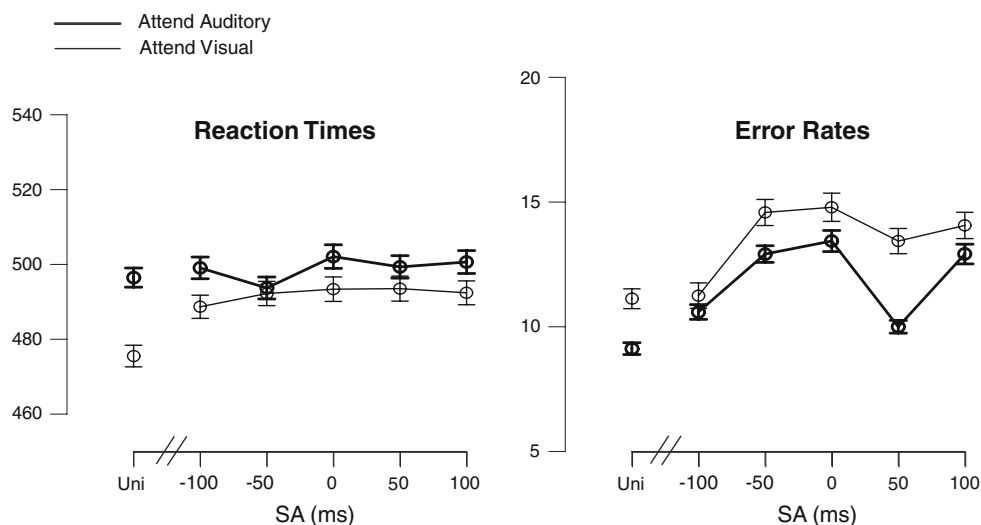


Fig. 3 Mean response times and error rates to unisensory (Uni = Unisensory target) and multisensory target stimuli. The other labels ‘-100’, ‘-50’, ‘0’, ‘+50’, ‘+100’ represent the multisensory trials, in which there are targets in the relevant modality and refer to the mean SOA (± 25 ms) of the accessory other-modality input of the multisensory stimulus. Thus, for the *attend-auditory* condition, the

with the mean RTs for the unisensory control stimulus targets. For the multisensory stimuli considered alone, no significant effect of the factor SOA on the RTs ($F(4,16) < 1$) was found. Although Fig. 3 indicates that responses to auditory targets were somewhat slower than those to visual targets, no significant main effects of Attention were found ($F(1,19) < 1$). Pair-wise comparisons confirmed that the responses to unisensory stimuli were faster than those to the multisensory stimuli ($p < 0.05$).

Error rates

Figure 3b shows the mean error rates for the target stimuli in each of the SOA conditions, along with the mean error rates for the unisensory control stimulus targets. For the multisensory targets, SOA significantly affected error rates ($F(4,16) = 4.43$; $p = 0.05$). This effect could be explained using a cubic polynomial fit ($F(1,19) = 10.7$; $p < 0.005$), confirming the pattern observed in Fig. 3b that for the multisensory stimuli the error rates were still relatively low when the unattended input was presented well before (i.e., at -100 ms), or immediately after (i.e., at +50 ms) the attended one, but increased at other SOAs.

False alarm rates

False alarm rates were computed for each condition separately and submitted to a MANOVA with Attention (attend-visual or attend-auditory) as within-subject factor. These tests indicated that false alarm rates did not differ

auditory input of the multisensory AV stimulus was the target, and the visual input the standard, occurring at the relative onset time indicated. Accordingly, for the *attend-visual* condition, the visual input of the multisensory AV stimulus was the target, with an auditory standard occurring at the relative onset time indicated

significantly between the attend-visual and attend-auditory conditions ($F(1,19) < 1$). Overall false alarm rates were about 5%.

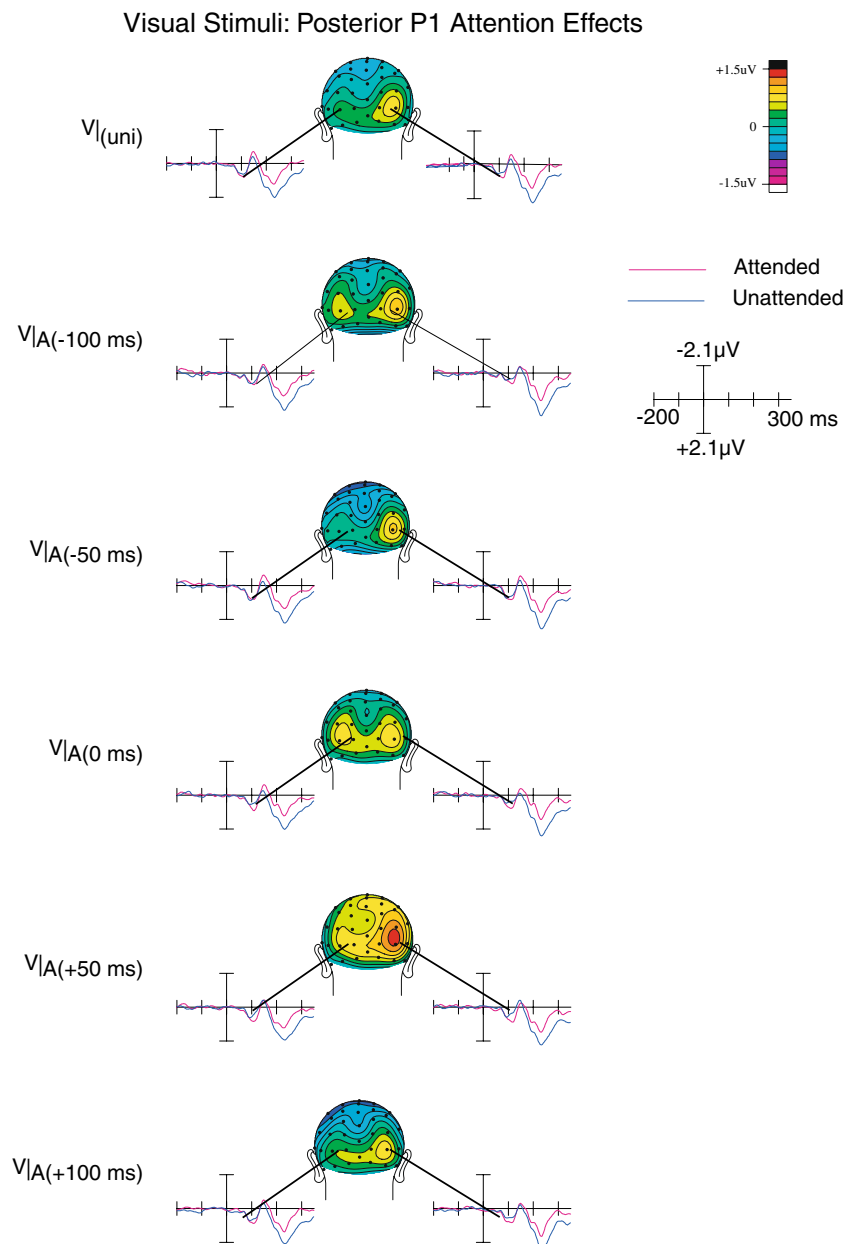
Event-related potentials

As noted above, the focus of this paper is on the ERPs to the visual inputs of the multisensory stimuli and on the interactions of intermodal attention and audio-visual asynchrony on these responses. In these extracted ERPs, attention effects were identified on four different components. Of these four components, the attention effects on the posterior P1 and on an anterior and posterior N1 component were modulated by SOA. In contrast, the fourth effect, the occipital selection negativity, remained constant across SOA.

Posterior P1

As shown in Fig. 4, the posterior P1 attention effect peaking at around 100 ms was characterized by a predominantly right hemispheric topographical distribution, with a strong focus around parietal–occipital electrodes P3i and P4i. The P1 components that were elicited by the visual inputs of the multisensory stimuli were larger when they were attended, compared to when they were unattended ($F(1,19) = 6.1$; $p < 0.05$). Although we did not find an interaction between Attention and SOA, a significant three-way interaction between Attention, SOA, and Laterality ($F(4,16) = 3.24$; $p < 0.05$) was observed. This interaction confirms the observation that the P1 attention effect differed across

Fig. 4 Occipital P1 attention effects on the visual ERPs. An early P1 component (~ 110 ms after onset of the visual stimulus) could be observed for the visual ERPs extracted from the multisensory trials, which was lateralized predominantly over the right hemisphere. This component was the largest when the auditory and visual stimuli were presented at the point of near optimal subjective alignment. It was also present for the unisensory visual control stimuli



SOAs, and also more strongly over the right hemisphere. Although Fig. 4 suggests that a P1 amplitude effect was also present for the unisensory stimuli, no statistical evidence could be obtained for this observation ($F(1,19) = 1.4$; $p < 0.25$).

Anterior N1

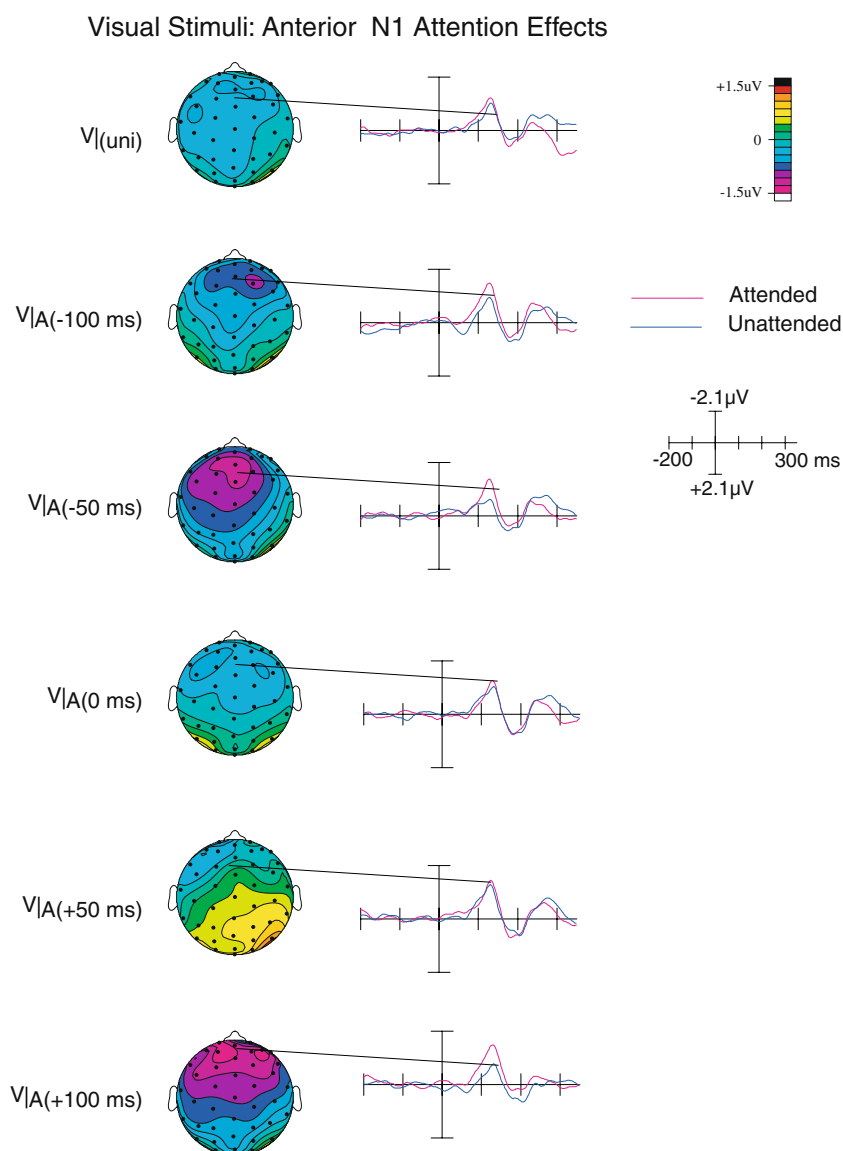
Shortly after the latency of the posterior P1, the anterior N1 component (peaking at around 140 ms) could be observed over the anterior scalp areas (see Fig. 5). This component was larger (more negative) when the visual inputs were attended than when they were unattended ($F(1,19) = 8.70$; $p < 0.01$). In addition, there was also a significant interac-

tion between Attention and SOA ($F(4,16) = 3.11$; $p < 0.05$). In contrast, no significant attention effect was observed on this component for the unisensory visual stimuli, suggesting that the anterior N1 attention effect here was specific to the processing of multisensory stimuli. Moreover, since the magnitude of this attention effect varied with SOA, we suggest that this effect may reflect processes related to resolving the temporal misalignment of audiovisual stimuli.

Posterior N1

The posterior N1, peaking at around 150–180 ms post-visual-stimulus, was significantly larger (i.e., more negative) when the visual stimuli were attended compared to

Fig. 5 Anterior N1 attention effects on the visual ERPs. Starting at around the same latency, an anterior N1 attention effect was observed, that was present only on multisensory stimuli, in particular at those SOAs, where the visual and auditory inputs were suboptimally aligned

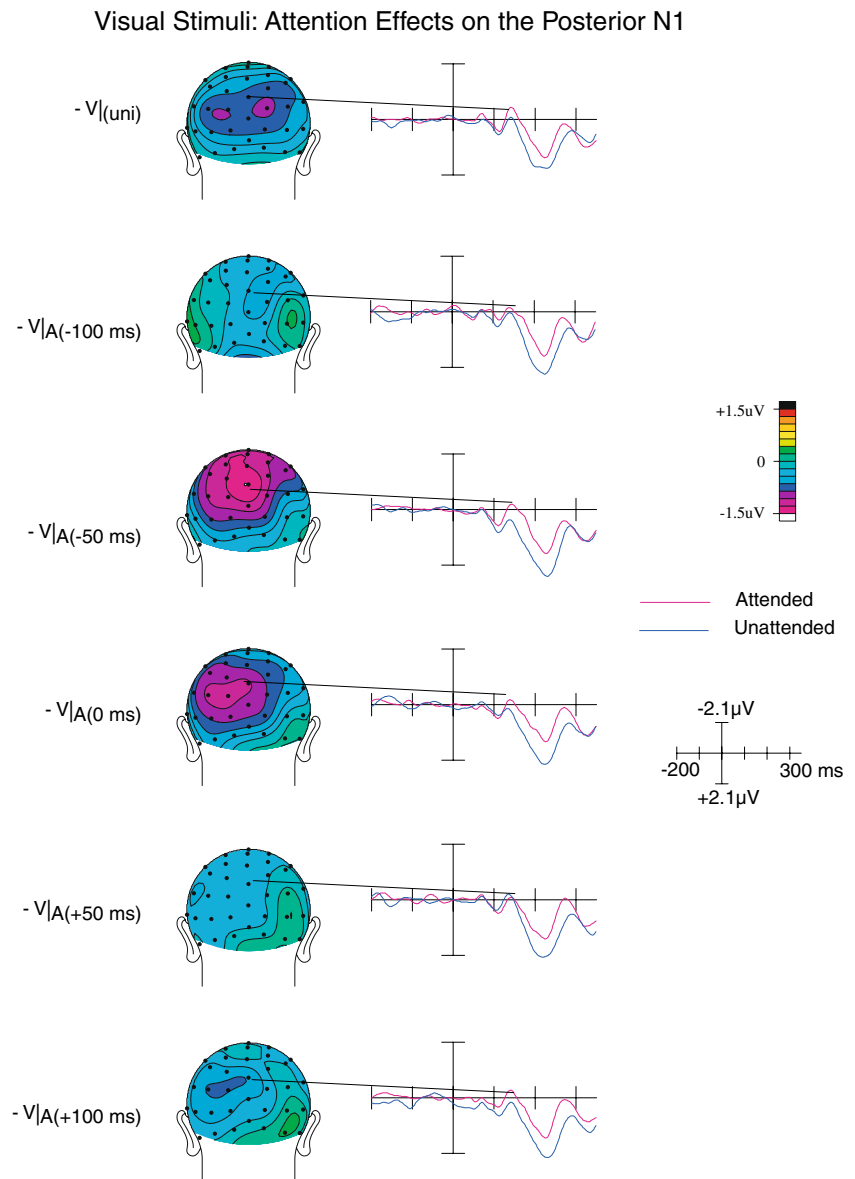


when they were unattended, as shown by a main effect of Attention on this component ($F(1,19) = 6.24$; $p < 0.05$; see Fig. 6). Although overall effect of SOA was not observed ($F(4,16) = 1.77$; $p > 0.1$), we did observe a significant *interaction* between Attention and SOA ($F(4,16) = 3.79$; $p < 0.05$), which appeared to result from this attention effect being largest when the auditory stimulus occurred either slightly before or synchronous with the visual stimulus (relative timing of -50 or 0 ms). A significant attention effect was also observed on the N1 component elicited by unisensory visual stimuli ($F(1,19) = 6.15$; $p < 0.05$).

Upon closer inspection, scalp topography of this attention effect appeared to differ between the unisensory and multisensory trial types. More specifically, whereas the topography of the unisensory N1 effect was characterized by a bilateral posterior scalp distribution, that of the multisensory N1 effect was characterized by a more mediopari-

etal, although perhaps somewhat left-sided distribution, particularly present when the auditory stimuli occurred just before or concurrent with the visual component. To test whether these topographies did indeed differ, all mean amplitudes for the attended-minus-unattended difference wave of the $V|_{uni}$, $V|_{A(-50\text{ ms})}$, and $V|_{A(0\text{ ms})}$ (where these effects were observed to be the most pronounced) were subjected to a vector normalization procedure (McCarthy and Wood 1985). The resulting normalized values were compared against each other using an ANOVA containing the within-subjects factors Stimulus Type (two levels) and Channel (64 levels). The presence of a significant interaction between Stimulus Type and Channel confirmed that the scalp topography of the visual N1 attention effect for the multisensory stimulus types differed from the topography of the N1 effect on the unisensory control stimulus ($F(63,1197) = 2.00$; $p < 0.05$; $\epsilon = 0.1$), suggesting that

Fig. 6 Posterior N1 attention effects on the visual ERPs. At around the latency of the posterior N1, differences between attended and unattended visual ERPs extracted from the multi-sensory trials were maximal when the auditory stimulus was presented just prior to, or near simultaneously with, the evoking visual stimulus. At the point of optimal temporal alignment, the N1 attention effect is attenuated



either additional neural generators, or a different configuration of neural generators, are active when an auditory input is presented immediately prior to a visual input.

Occipital selection negativity

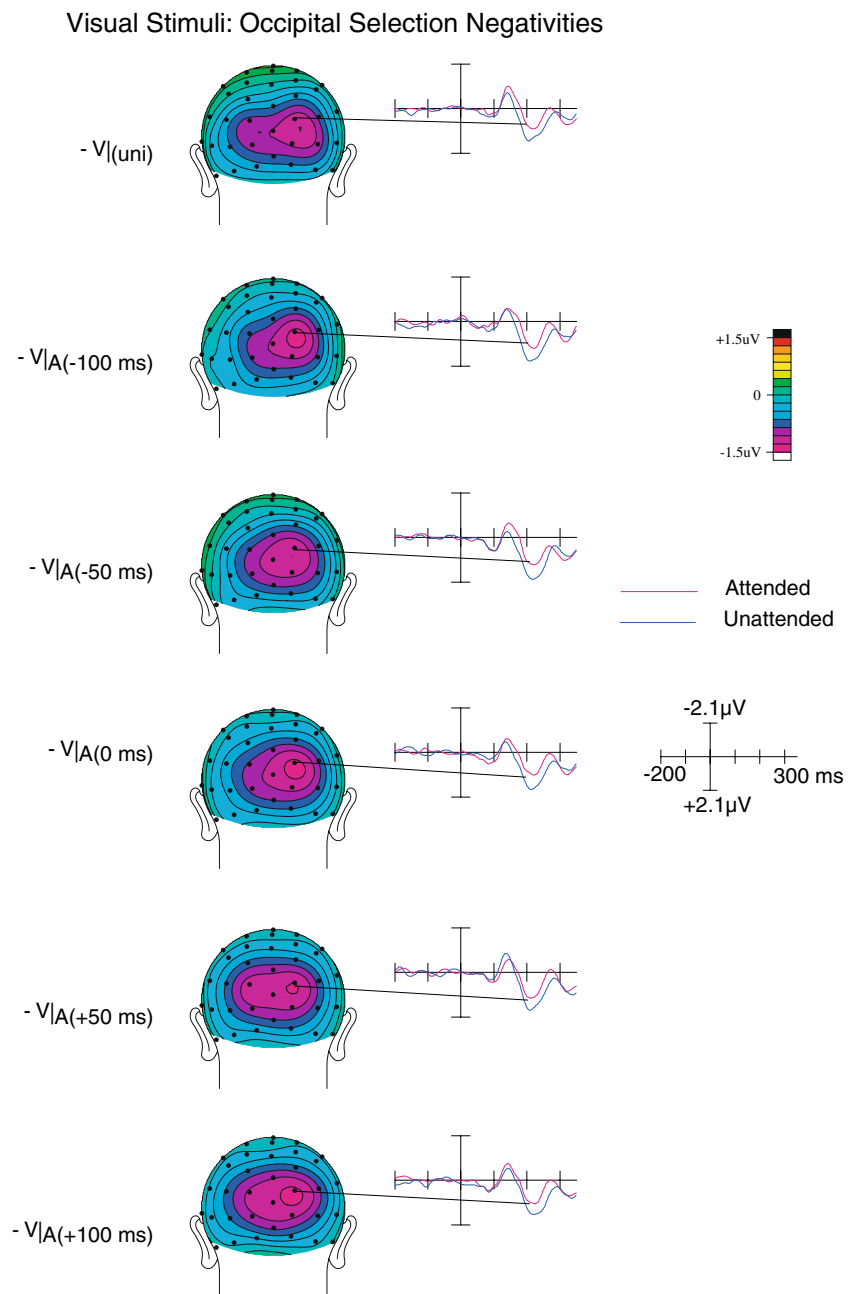
Beginning at around 200 ms, ERPs elicited by attended visual inputs became significantly more negative than those elicited by unattended visual inputs (see Fig. 7). This difference, which appeared to be similar to the slow, long-latency posterior attention effect known as the occipital selection negativity, was reflected by the presence of a significant main effect of Attention ($F(1,19) = 10.8$; $p < 0.005$). For this activity, although we observed a main effect of SOA ($F(4,16) = 3.25$; $p < 0.05$), there was no interaction between Attention and SOA, indicating that the attention-

related occipital selection negativity was present for all SOAs and did *not* significantly differ between SOA conditions, suggesting that by this time a stable representation had been established that was beyond influence by the asynchrony between the audio and visual inputs. Finally, it should be noted that the occipital selection negativity attention effect was also present for unisensory stimuli ($F(1,19) = 4.64$; $p < 0.05$).

Discussion

The main goal of this study was to investigate the impact of intermodal attention on the neural processes involved in multisensory processing, especially those that interact with the degree of onset asynchrony between the auditory and

Fig. 7 Occipital selection negativity. A relatively long-latency visual attention effect, the occipital selection negativity was found to be significant across all SOAs for the visual component of the multisensory trials, and did not vary as a function of SOA. It was also present on the unisensory stimuli



visual components of a multisensory stimulus. Significant interactions between attention and stimulus asynchrony were observed. In particular, attention effects on the visually elicited P1 and N1 components differed as a function of audiovisual SOA, that is, attention effects on the right-hemisphere visual P1 were larger when the auditory and visual inputs were optimally aligned, whereas the attention effects on the visual N1 component were relatively small when this was the case. In contrast, the longer-latency occipital selection negativity attention effect was not influenced by SOA. In addition, we observed an anterior N1 attention effect component that was only present on the visual ERPs elicited by the multisensory stimuli, and which

was most pronounced when the auditory and visual stimuli were *suboptimally* aligned.

Temporal asynchrony and multisensory processing

SOA effects in behavioral data

Overall, response times to multisensory targets were longer than those to unisensory targets. This finding stands somewhat in contrast to the behavioral advantages associated with multisensory targets that have been often reported in the literature. However, these advantages have typically been found when the multisensory targets are characterized

by redundant target features (Giard and Peronnét 1999; Molholm et al. 2002; Senkowski et al. 2006). In contrast, when the target features are non-redundant (Fort et al. 2002; Gondan et al. 2005), such behavioral advantages do not seem to occur. In the present study the target stimuli contained only the target feature in one modality, and the near-simultaneous presentation of a non-target stimulus in another modality, thus would not have been helpful in detecting the target in the attended modality and might have actually interfered with the target processing in the relevant modality. This might have in turn led to the need for invocation of an attentional selection mechanism. Thus, a response conflict between an attended target stimulus paired with an unattended non-target stimulus might have resulted in the slight increase in response times.

A similar conclusion can be drawn from the accuracy data, which show that targets that were part of a multisensory stimulus were generally processed with slightly lesser accuracy. In addition, as shown in Fig. 3, error rates generally increased as the audiovisual SOA increased. More specifically, when the other-modality stimulus was presented well before the relevant target, only small increases in error rate were observed, until the relevant target and the irrelevant other-modality non-target were presented near simultaneously, suggesting an increase in the competition for attentional capacity up to the point of near simultaneity. One notable exception to this pattern is the +50 SOA, where error rates were lower. Although it is not clear what might be the explanation for this effect, one possibility could be that at these relatively short latencies the second stimulus in the sequence was presented late enough, such that it had relatively little impact on ongoing perceptual processes (such as those involved in determining whether multisensory stimuli should be integrated or not), but early enough not to interfere with higher-order cognitive processes, such as cross modal distraction (e.g., Escera et al. 2000). To sum up, the results from the analysis of reaction time and error rates show that both SOA and intermodal attention influence behavioral performance.

Interactions between SOA and attention effects in the ERPs

The ERP data also underscored the behavioral observation that the auditory and visual stimulus inputs competed for attentional capacity. The attention effect on the posterior P1 component was present on unisensory as well as on multisensory trials, and can therefore be considered as a generic effect of intermodal attention, although the latency of this effect was somewhat earlier than the typical effects of non-spatial intermodal attention that have been previously reported (Talsma and Kok 2001). Interestingly, this attention effect was largest when the auditory input occurred around 50 ms after the visual input, an asynchrony that has

been proposed to be the optimum offset for the subjective perception of simultaneity (Lewald and Guski 2003), leading to greater multisensory integration. Thus, this intermodal attention-related P1 enhancement could reflect a low-level mechanism that boosts visual selection processing when auditory stimuli are particularly close to the point that yields subjective simultaneity.

The anterior N1 attention effect was unique to the multisensory stimuli and was not observed for the unisensory stimuli. Moreover, since this attention effect was observed predominantly when the temporal offset between visual and auditory inputs was *suboptimal*, we speculate that this effect could reflect a process that is involved in determining whether or not a visual input should or could be temporally aligned to the auditory inputs (Shams et al 2001). Since the temporal resolution of the visual modality is lower than that of the auditory modality (e.g., DiLuca et al. 2009), might be the demand on processes involved in temporal realignment of visual stimuli to the auditory stimuli is larger when the visual stimuli are attended than when they are not attended, giving rise to the pattern observed. Considering that attending to visual stimuli serves the purposes of enhancing stimulus clarity, and reducing ambiguity, it is plausible that when a visual stimulus becomes relevant, the attentional resolving mechanisms seek to enhance all aspects of the visual stimulus, including its location, visual features, and its temporal characteristics. In line with this, Correa and Nobre (2008) have reported evidence that attention can enhance the temporal acuity of unisensory stimuli. Thus, the need for the realignment of a visual stimulus to an auditory one would be particularly large when the visual stimulus is attended. Therefore, in contrast to the occipital P1, which presumably reflects a boost in visual saliency, the N1 seems likely to reflect a somewhat later stage of multisensory integration.

Interestingly, we found that the early posterior unisensory N1 component was larger for attended stimuli than for unattended stimuli. Since the N1 amplitude is typically not modulated by non-spatial forms of attentional feature selection (e.g., Kenemans et al. 1993), our findings suggest that, unlike intramodal non-spatial selective attention processes, the intermodal attention to a centrally presented stimulus may indeed exert a modulatory influence on early sensory ERPs (<200 ms). Notably, the N1 amplitude differences between attended and unattended stimuli (i.e., the attention effects) were smallest when the auditory stimulus was presented near simultaneously (effects on the anterior N1) and when the auditory stimulus followed the visual input by about 50 ms (effects on the anterior and posterior N1). Previous research has shown that a delay of about 50 ms of the auditory stimulus relative to the visual onset is optimal for multisensory integration (e.g., Lewald and Guski 2003). Therefore, a reduction in N1 differences between attended

and unattended visual stimulus inputs of the multisensory stimuli at these latencies would be consistent with a multisensory integration account. More specifically, even though attention would not be directed to the auditory stimuli, the optimal temporal alignment may enable the auditory and visual inputs to be more effectively integrated into a multisensory representation that would be encompassed entirely by attention.

Thus, the above observations may therefore be an indicative of “spreading of attention” across modalities analogous to effects previously reported (Busse et al. 2005; Talsma et al. 2007). In the Busse et al. (2005) study, ERP responses to auditory stimuli were characterized by a prolonged late negative component when these stimuli were paired with an attended visual stimulus (and even when it was coming from a different spatial location), which was not present when the same stimulus was paired with an unattended visual stimulus. A similar effect, suggested to be the multisensory counterpart of the auditory “processing negativity”, was reported on the auditory input of a multisensory stimulus when only the visual modality was attended (Talsma et al. 2007). In conclusion, we interpret the occipital N1 component as being related to the initial phases of a spread of attention from the visual to the auditory modality.

Note that these effects on the occipital N1 are somewhat different from those on the P1. Whereas the attention effects on the P1 were largest when the auditory and visual stimuli are optimally aligned, attention effects on the occipital N1 are diminished for optimally aligned auditory and visual inputs. This difference suggests that a possible initial boost of processing in the already attended modality is followed by a spread of attention to the unattended modality, in particular when auditory stimuli are optimally aligned.

Finally, a stable later effect of attention was found, starting at 200 ms and continuing beyond 300 ms that was identified as the occipital selection negativity. This attention effect is consistent with previously reported similar modulations due to non-spatial intermodal attention (Alho et al. 1992; De Ruiter et al. 1998; Talsma and Kok 2001). In addition, Woods et al. (1992) observed a similar negative displacement related to intermodal attention using lateralized stimuli. Thus, the effects of intermodal attention in the current study appear to extend these results to the processing during near-simultaneous multisensory stimuli as well. Interestingly, the occipital selection negativity was the only attention component that did *not* interact with SOA of the multisensory stimulus inputs. This finding is consistent with previous interpretations of the selection negativity as reflecting a cognitive evaluation process that occurs after attentional selection (and in this case multisensory integration) would have already taken place (Smid et al. 1999). Thus, this result suggests that, by the time of occurrence of

the occipital selection negativity, the temporal realignment and multisensory integration due to the adjacent auditory stimulus component has already occurred. This in turn suggests that the visual representation is stable by this time, and receives attentionally enhanced, higher-level processing that is independent of the relative timing of the auditory stimulus. Thus, the present study underscores the dissociation between an early (semi-) automatic multisensory analysis process and later stages of higher-level processing.

A similar distinction of levels of multisensory processing was reached by Magnée et al. (2008a, b), who demonstrated that in a group of individuals diagnosed with pervasive developmental disorders (DPP) the early stages of multisensory integration were still intact, whereas the later stages of multisensory integration were impaired. Since disrupted multisensory processing is considered to be a part of various clinical populations, including autism spectrum disorder (Iarocci and McDonald 2006) or Schizophrenia (Ross et al. 2007), the present paradigm could be a viable tool to study the interactions between multisensory integration and cognition in these patient groups.

Summary and conclusions

This study revealed behavioral and electrophysiological evidence for the influence of attention on resolving the input timing differences in a multisensory stimulus. We found that ERPs that were time-locked to the visual component of the audiovisual stimuli showed several attention effects. Of these, the amplitude of the attention effects on early latency components (posterior P1, anterior N1, and posterior N1) were modulated by SOA. Furthermore, attention effects on the posterior P1 and N1 components also occurred for unisensory stimuli. These P1 and N1 effects are therefore likely to be multisensory extensions of basic attention processes. In contrast, we observed a frontal N1 attention effect that was present only on the multisensory stimuli, and, particularly so, when the temporal alignment of the auditory and visual inputs was less than optimal. We therefore suggest that the frontal N1 effect might reflect the processing or adjusting for the temporal alignment of visual inputs to the auditory inputs. In contrast, a longer-latency attention effect, the occipital selection negativity, was not influenced by SOA. We conclude from these findings that even small temporal asynchronies, occurring within the temporal window of integration, are detected at early latencies, and that somewhat later in time frontal brain areas contribute to the processing and resolving of these temporal asynchronies. In contrast, a longer-latency attention effect, the occipital selection negativity, was observed for all SOAs and was not influenced by SOA, suggesting that by this time a stable multisensory representation had been established that was

beyond influence from the within-integration-window asynchronies between the audio and visual inputs.

Acknowledgments We wish to thank Tineke Grent-‘t-Jong, Ken Roberts, and Chad Hazlett for technical support. Furthermore, we would like to thank Tracy Doty and Vince Wu for their help during data collection. This study was supported by grants from NIH (R01 NS051048) and from NSF (BCS-05-24031) to Marty G. Woldorff, and by a grant from the German Research Foundation (SE 1859/1-1) to Daniel Senkowski.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Alho K, Woods DL, Algazi A, Näätänen R (1992) Intermodal selective attention II: effects of attentional load on processing auditory and visual stimuli in central space. *Electroencephalogr Clin Neurophysiol* 82:356–368
- Andersen TS, Mamassian P (2008) Audiovisual integration of stimulus transients. *Vis Res* 48:2537–2544
- Burock MA, Buckner RL, Woldorff MG, Rosen BR, Dale AM (1998) Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *NeuroReport* 9:3735–3739
- Busse L, Woldorff MG (2003) Implications of the ERP omitted stimulus response to “no-stim” events in fast-rate event-related fMRI designs. *Neuroimage* 18:856–864
- Busse L, Roberts KC, Crist RE, Weissman DH, Woldorff MG (2005) The spread of attention across modalities and space in a multisensory object. *Proc Natl Acad Sci USA* 102:18751–18756
- Correa A, Nobre AC (2008) Spatial and temporal acuity of visual perception can be enhanced selectively by attentional set. *Exp Brain Res* 189:339–344
- De Ruiter MB, Kok A, Van der Schoot M (1998) Effects of inter- and intramodal selective attention to non-spatial stimuli: an event-related potential analysis. *Biol Psychol* 49:269–294
- DiLuca M, Machulla T-K, Ernst MO (2009) Recalibration of multisensory simultaneity: crossmodal transfer coincides with a change in perceptual latency. *J Vis* (submitted)
- Escera C, Alho K, Schröger E, Winkler I (2000) Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiol Neurootol* 5:151–166
- Fendrich R, Corballis PM (2001) The temporal cross-capture of audition and vision. *Percept Psychophys* 63:719–725
- Fort A, Delpuech C, Pemier J, Giard MH (2002) Early auditory-visual interactions in human cortex during nonredundant target identification. *Cogn Brain Res* 14:20–30
- Giard MH, Peronnét F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* 11:473–490
- Gondan M, Röder B (2006) A new method for detecting interactions between the senses in event-related potentials. *Brain Res* 1073:389–397
- Gondan M, Niederhaus B, Rösler F, Röder B (2005) Multisensory processing in the redundant-target effect: a behavioral and event-related potential study. *Percept Psychophys* 67:713–726
- Iarocci G, McDonald JJ (2006) Sensory integration and the perceptual experience of persons with autism. *J Autism Dev Disord* 36:77–90
- Keetels M, Vroomen J (2005) The role of spatial disparity and hemifields in audio-visual temporal order judgments. *Exp Brain Res* 167:635–640
- Kenemans JL, Molenaar PCM, Verbaten MN (1991) Models for estimation and removal of artifacts in biological signals. In: Weitkunat R (ed) *Digital biosignal processing*. Elsevier, New York
- Kenemans JL, Kok A, Smulders FTY (1993) Event-related potentials to conjunctions of spatial-frequency and orientation as a function of stimulus parameters and response requirements. *Electroencephalogr Clin Neurophysiol* 88:51–63
- King AJ (2005) Multisensory integration: strategies for synchronization. *Curr Biol* 15:R339–R341
- Lewald J, Guski R (2003) Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cogn Brain Res* 16:468–478
- Lewald J, Guski R (2004) Auditory-visual temporal integration as a function of distance: no compensation for sound-transmission time in human perception. *Neurosci Lett* 357:119–122
- Lewald J, Ehrenstein WH, Guski R (2001) Spatio-temporal constraints for auditory-visual integration. *Behav Brain Res* 121:69–79
- Magnée MJCM, de Gelder B, van Engeland H, Kemner C (2008a) Audiovisual speech integration in pervasive development disorder: evidence from event-related potentials. *J Child Psychol Psychiatry* 49:995–1000
- Magnée MJCM, de Gelder B, van Engeland H, Kemner C (2008b) Atypical processing of fearful face-voice pairs in pervasive developmental disorder: an ERP study. *Clin Neurophysiol* 119:2004–2010
- McDonald JJ, Teder-Sälejärvi WA, Di Russo F, Hillyard SA (2005) Neural basis of auditory-induced shifts in visual time-order perception. *Nat Neurosci* 8:1197–1202
- Meredith MA (2002) On the neuronal basis for multisensory convergence: a brief overview. *Brain Res Cogn Brain Res* 14:31–40
- Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. 1. Temporal factors. *J Neurosci* 7:3215–3229
- 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
- Morein-Zamir S, Soto-Faraço S, Kingstone A (2003) Auditory capture of vision: examining temporal ventriloquism. *Cogn Brain Res* 17:154–163
- Näätänen R (1992) *Attention and brain function*. Lawrence Erlbaum, Hillsdale, NJ
- Navarra J, Vatakis A, Zampini M, Soto-Faraço S, Humphreys W, Spence C (2005) Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration. *Cogn Brain Res* 25:499–507
- Ross LA, Saint-Amour D, Leavitt VM, Molholm S, Javitt DC, Foxe JJ (2007) Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophr Res* 97:173–183
- 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, 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, Talsma D, Grigutsch M, Herrmann CS, Woldorff MG (2007a) Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia* 45(3):561–571
- Senkowski D, Saint-Amour D, Kelly SP, Foxe JJ (2007b) Multisensory processing of naturalistic objects in motion: a high-density

- electrical mapping and source estimation study. *Neuroimage* 36:877–888
- Shams L, Kamitani Y, Thompson S, Shimojo S (2001) Sound alters visual evoked potentials in humans. *NeuroReport* 12:3849–3852
- Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. *Cogn Brain Res* 14:147–152
- Smid HGOM, Jakob A, Heinze HJ (1999) An event-related brain potential study of visual selective attention to conjunctions of color and shape. *Psychophysiology* 36:264–279
- Spence C, Squire S (2003) Multisensory integration: maintaining the perception of synchrony. *Curr Biol* 12:R519–R521
- Stein BE, Meredith MA (1993) *The merging of the senses*. MIT press, Cambridge, MA
- Stekelenburg JJ, Vroomen J (2005) An event-related potential investigation of the time-course of temporal ventriloquism. *NeuroReport* 16:641–644
- Talsma D (2008) Auto-adaptive averaging: detecting artifacts in event-related potential data using a fully automated procedure. *Psychophysiology* 45:216–228
- Talsma D, Kok A (2001) Nonspatial intermodal selective attention is mediated by sensory brain areas: evidence from event-related potentials. *Psychophysiology* 38:736–751
- Talsma D, Woldorff MG (2005a) Attention and multisensory integration: multiple phases of effects on the evoked brain activity. *J Cogn Neurosci* 17:1098–1114
- Talsma D, Woldorff MG (2005b) Methods for the estimation and removal of artifacts and overlap in ERP waveforms. In: Handy TC (ed) *Event-related potentials: a methods handbook*. MIT Press, Cambridge, MA, pp 115–148
- Talsma D, Doty TJ, Woldorff MG (2007) Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb Cortex* 17:691–701
- 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. *Cogn Brain Res* 14:106–114
- Vroomen J, De Gelder B (2004) Temporal ventriloquism: sound modulates the flash-lag effect. *J Exp Psychol Hum Percept Perform* 30:513–518
- Vroomen J, Keetels M (2006) The spatial constraint in intersensory pairing: no role in temporal ventriloquism. *J Exp Psychol Hum Percept Perform* 32:1063–1071
- Wallace MT, Meredith MA, Stein BE (1993) Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. *J Neurophysiol* 69:1797–1809
- Woldorff MG (1993) Distortion of ERP averages due to overlap from temporally adjacent ERP: analysis and correction. *Psychophysiology* 30:98–119
- Woods DL, Alho K, Algazi A (1992) Intermodal selective attention I: effects of event-related potentials to lateralized auditory and visual stimuli. *Electroencephalogr Clin Neurophysiol* 82:341–355