

The electrophysiological time course of the interaction of stimulus conflict and the multisensory spread of attention

U. Zimmer,¹ S. Itthipanyanan,¹ T. Grent-'t-Jong^{1,2} and M.G. Woldorff^{1,3}

¹Center for Cognitive Neuroscience, Duke University, LSRC Rm B242, Box 90999, Durham, NC, USA

²Departments of Experimental Psychology and Psychopharmacology, University of Utrecht, Utrecht, The Netherlands

³Department of Psychiatry, Duke University, Durham, NC, USA

Keywords: auditory, EEG, ERP, visual

Abstract

Previously, we have shown that spatial attention to a visual stimulus can spread across both space and modality to a synchronously presented but task-irrelevant sound arising from a different location, reflected by a late-onsetting, sustained, negative-polarity event-related potential (ERP) wave over frontal–central scalp sites, probably originating in part from the auditory cortices. Here we explore the influence of cross-modal conflict on the amplitude and temporal dynamics of this multisensory spreading-of-attention activity. Subjects attended selectively to one of two concurrently presented lateral visually-presented letter streams to perform a sequential comparison task, while ignoring task-irrelevant, centrally presented spoken letters that could occur synchronously with either the attended or unattended lateral visual letters and could be either congruent or incongruent with them. Extracted auditory ERPs revealed that, collapsed across congruency conditions, attentional spreading across modalities started at approximately 220 ms, replicating our earlier findings. The interaction between attentional spreading and conflict occurred beginning at approximately 300 ms, with attentional-spreading activity being larger for incongruent trials. Thus, the increased processing of an incongruent, task-irrelevant sound in a multisensory stimulation appeared to occur some time after attention has spread from the attended visual part to the ignored auditory part, presumably reflecting the conflict detection and associated attentional capture requiring accrual of some multisensory interaction processes at a higher-level semantic processing stage.

Introduction

While driving a car, maintaining visual attention on the surrounding traffic is essential, and ignoring irrelevant auditory distraction generally beneficial. However, even though ignoring irrelevant distracting conversation from a passenger might improve the effectiveness of our driving, ignoring his or her warning about an upcoming, potentially dangerous, event could be quite disadvantageous. Evolutionarily speaking, it seems beneficial that our brains are capable of parsing and assessing conflicting, possibly relevant, information from concurrently occurring input.

Recently, we showed that attention can spread from an attended visual stimulus to a task-irrelevant, simultaneously presented, auditory stimulus, even when the two arise from different spatial locations (Busse *et al.*, 2005). The spreading of attention was manifested as a late-onsetting (220 ms), sustained, frontally distributed, event-related potential (ERP) wave elicited by auditory stimuli that occurred synchronously with an attended visual stimulus, relative to when they occurred with an unattended one. In that study, however, the effects were observed using very simple visual and auditory stimuli with no

apparent higher-level representation-related congruence or incongruence.

The way the brain responds to conflicting stimulus input has been studied extensively using classic paradigms such as the Stroop (1935) and flanker (Eriksen & Eriksen, 1974) paradigms. These studies have typically observed an increased negative-polarity ERP wave to incongruent (i.e. conflicting) vs. congruent stimulus input over frontal–central scalp sites, peaking 250–400 ms poststimulus (e.g. West & Alain, 1999; Yeung *et al.*, 2004; Bartholow *et al.*, 2005; Wendt *et al.*, 2007; Appelbaum *et al.*, 2009), an electrophysiological response thought to arise in part from increased activity in the anterior cingulate cortex (Van Veen & Carter, 2002; Fan *et al.*, 2003, 2007; Hanslmayr *et al.*, 2008). Most such stimulus-conflict studies, however, have been carried out within a single modality. Moreover, studies that have investigated cross-modal conflict have typically had the auditory and visual sensory components occurring at the same location (e.g., Van Atteveldt *et al.*, 2007; Fiebelkorn *et al.*, 2010). Thus, it is not at all clear how the spreading of attention across the spatially separated sensory components of a multisensory stimulus might vary as a function of whether, and how, those components conflict.

Here, we used the high temporal resolution of ERP recordings to examine the influence of cross-modal conflict on the amplitude and timing of multisensory spreading-of-attention activity. We presented letter–sound combinations while subjects selectively attended to one

Correspondence: Dr Ulrike Zimmer, ¹Center for Cognitive Neuroscience, as above.
E-mail: ulrike.zimmer@duke.edu

Received 1 October 2009, revised 3 March 2010, accepted 11 March 2010

of two lateralized visually-presented letter streams and ignored task-irrelevant, centrally presented, spoken letters that could occur synchronously with either the attended or unattended lateral visual letters and could be either congruent or incongruent with them. We hypothesized, first of all, that spreading-of-attention activity would probably differ in amplitude as a function of cross-modal incongruence, either being smaller due to suppression of the conflicting auditory input or larger due to increased attentional capture by the conflicting input. Secondly, the influence of cross-modal incongruence might occur in close temporal conjunction with the attentional spreading, onsetting at the same time, or it could be contingent upon the accrual of some attentional spreading and multisensory interaction processes, thus onsetting sometime after.

Material and methods

Participants

Twenty-six healthy, right-handed participants (aged 18–35 years; equal numbers of males and females) participated in the experiment. Ten of these participants were excluded from the final data analysis because > 50% of their trials had to be rejected due to eye-movement

artifacts. All 26 participants gave written informed consent after receiving an explanation of the procedures, using a protocol approved by the Duke University Institutional Review Board. Participants were paid \$15 per hour for their participation.

Paradigm

The study incorporated a variation of the classic ‘one-back’ paradigm, consisting of a stream of compound trials made up of sequential pairs of visual letters (‘A’, ‘X’ or ‘H’) that could be either identical or different and could be accompanied by a task-irrelevant, simultaneously presented, spoken letter of the same or different identity. All visual stimuli in these compound trials were presented randomly from a left or right lower visual field location, whereas the occasionally accompanying auditory stimulus was always presented from a central position (see Fig. 1).

For each run, participants were instructed to covertly attend to one of the two visual streams (left or right location) and to ignore both the visual stimuli on the other side and all of the centrally presented spoken letters sounds. Visual stimuli consisted of sequentially presented pairs of letter stimuli (A or X only) at one of the two (attended or unattended) lower visual field locations; pairs of letters

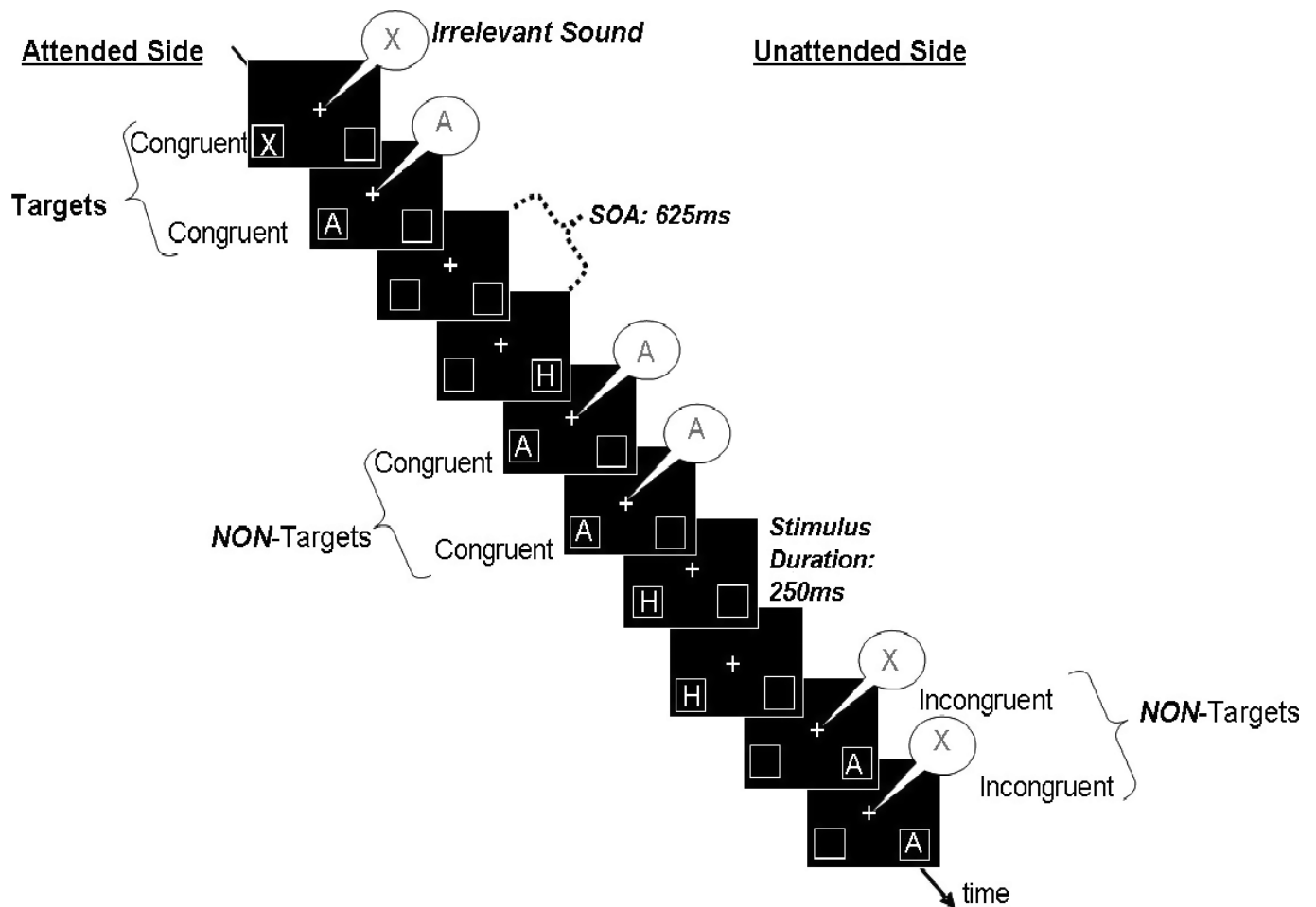


FIG. 1. Task paradigm, shown for runs with subjects’ attention to the left side. Two visual-letter streams were presented randomly to the left and right visual fields, including two task-relevant letters (A and X). The task of the subject was to fixate on the central white cross, attend to the stream on the designated side, and press a button for a switch of task-relevant visual letters in a sequential pair on that side. To reduce button-press-related activity, targets requiring a button press comprised only 15% of the trials, allowing for analysis of attended vs. unattended nontarget responses, free of target-related and motor-related processing. Two-thirds of the lateral visual letters (both attended and unattended) were accompanied by simultaneous, centrally presented, spoken letter sounds, 50% of which were incongruent and 50% congruent with the visual letter. Importantly, subjects were instructed to ignore the central spoken letters. SOA, stimulus onset separation.

were usually identical ('matched') but sometimes were not. At the attended location only, nonidentical pairs required responses of the subjects (i.e. these were thus target pairs). More specifically, subjects were instructed to press one of two buttons (with their right index finger) when the second letter of a covertly attended visual pair sequence was an A (and was thus preceded by an X), and press the other button when it was an X (and thus was preceded by an A).

Two-thirds of the letters of each of the lateralized visual letter-streams consisted of possible task-relevant (when in the attended stream) sequential letter pairs (X and/or A) and one-third were task-irrelevant pairs [H and/or no letter ('no-stim')], the latter serving to separate the task-relevant sequential pairs when they appeared in immediate succession within the same stream. These irrelevant pairs that served as separators consisted of either two Hs in a row, two 'no-stims' in a row, or an H followed by a 'no-stim' or *vice versa*, in equal proportion.

In the task-relevant sequential pairs (attended or unattended), 34% of the trials were visual-only trials (pure lateralized visual letters, no accompanying sound) and 66% were multisensory (lateralized visual letters accompanied by a synchronous, centrally presented, spoken letter A or X, male voice), which could be either congruent (50%) or incongruent (50%) with the corresponding lateralized visual letter in that trial. These task-relevant trials could contain the following types of sequential pairs: X-followed-by-X, A-followed-by-A, A-followed-by-X and X-followed-by-A. In these trials, 78% of the pairs included the visual sequentially matching letters AA and XX, which occurred in equal proportions. The remaining 22% of these trials consisted of the visual sequentially nonmatching letters AX and XA, which were the target sequence pairs to detect. These also occurred with equal probability.

The sequential pairs of visual-letter stimuli were presented in random order to the left and right side. In addition, whether the individual lateral visual stimulus pairs were accompanied by task-irrelevant central auditory letters was also randomized, such that the order of presentation of the multisensory context [pure (visual alone), congruent multisensory or incongruent multisensory] was unpredictable. All stimuli (both visual and auditory) were presented for a duration of 250 ms. The stimulus-onset-asynchrony (SOA) between the two stimuli of a sequential pair as well as between two sequential pairs was 625 ms.

The visual letters were presented 9.5° to the left or right of the fixation cross, and vertically 4° below fixation. They were presented in two rectangular boxes (size 3.8 × 5 cm²), which remained on the screen continuously during the entire runtime, serving as attentional anchors to assist the participant in maintaining a strong covert attentional focus at the spatial position of the to-be-attended visual letter stream.

All participants completed 14 runs, seven in which they attended covertly to the right lateral stream and seven in which they covertly attended to the left one (in randomized order). Each run lasted approximately 3.5 min, leading to a total experimental runtime of approximately 47 min.

EEG recording

The EEG was recorded from 128 channels mounted in a customized elastic electrode-cap (Duke128 Waveguard cap layout, made by Advanced Neuro Technology (ANT), the Netherlands), referenced to the average of all channels during recording. The 128 channels were equally spaced across the cap and covered the whole head from above the eyebrows to lower aspects of the occipital cortex (slightly past theinion). Electrode impedances were kept < 5kΩ for all 128 electrodes.

The ground electrode was placed on the collar bone. Horizontal eye movements were detected by two extra bipolar electrodes placed at the outer canthi of the eyes, whereas vertical eye movements or blinks were detected by another pair of bipolar electrodes, placed below and above the right eye. All EEG and EOG channels were recorded continuously in DC mode from a 128-channel, high-impedance ANT-Waveguard amplifier with active cable-shielding technology, which were digitized with a sampling rate of 512 Hz per channel for offline storage and analysis. Recording was done in a sound-attenuated, electrically shielded chamber, kept in relatively low lighting.

After the end of the 14 experimental runs, the locations of the electrodes were digitized for each participant using a 3-D spatial digitizer (Polhemus, Inc., USA). These locations were used later to more accurately calculate topographic distributions of the grand-average ERP waveforms based on the mean of all included participants' electrode locations (standard error of these locations across subjects and electrodes: *x*-axis, ±1.16 mm; *y*-axis, ±1.34 mm; *z*-axis, ±2.14 mm).

Data analysis

Behavioral data

The behavioral responses to the target stimuli (i.e. the second stimulus in sequentially presented pairs of nonmatching visual letters) were analyzed primarily to estimate the behavioral influence of conflicting sounds during visual stimulation (e.g. Van Atteveldt *et al.*, 2007). Only trials for which the behavioral responses occurred between 200 and 1000 ms after the second letter of a sequential target pair were considered for further behavioral analysis. Reaction times (RTs) and accuracy for correctly detected sequential letter pair orders were computed separately for the congruent, incongruent and pure-visual conditions. Repeated-measures ANOVAs were performed, using the within-subject factor Condition (congruent, incongruent, pure) to check for significant differences between conditions. Significance was inferred for Greenhouse–Geisser corrected *P*-values < 0.05.

EEG data

Data were preprocessed with the commercial software package ASA (Advanced Neuro Technology) and then analyzed further using custom ERSS software (Event-Related Potential Software System, UCSD, San Diego, CA, USA). Preprocessing included high-pass filtering of the data (Butterworth filter, cut-off frequency 0.016 Hz, linear roll-off 12 dB/oct) to exclude ultraslow DC drifts, prior to transformation of the data into ERSS format for further analysis.

We confined our EEG analyses to the first letter of the sequential pair, as these trials were not contaminated by button press-related or target detection-related processes. The continuous EEG data were divided into 800-ms epochs, time-locked to the first letter in the various pairs, and included a prestimulus baseline of 200 ms. Artifact rejection was performed off-line by discarding epochs contaminated by eye movements, eye blinks, excessive muscle activity, drifts, or amplifier blocking. Nonartifact EEG epochs were then averaged together, separately for the various trial types. These averages were then subsequently digitally low-pass filtered with a running-average filter of nine points (which at our sampling rate of 512 Hz corresponds to a low-pass filter cutoff of approximately 57 Hz) and re-referenced to the algebraic mean of the two mastoid electrodes. For the testing of multisensory spreading of neural activity, we used a frontal–central region of interest (ROI) that matched as closely as possible the four electrode positions where visual attention spreading was reported in Busse *et al.* (2005) (i.e. approximately Fz, FCz, FC1 and FC2 in the

standard 10–20 system, corresponding in our system to channels Z4, Z5, L4, and R4). To test for significant effects of the various conditions, statistical analyses were computed in successive time bins of 20 ms of averaged data, between 0 and 600 ms.

Check for effective covert visual spatial attention

To check whether the participants successfully maintained selectively directed covert attention to the designated letter stream, we examined the amplitude of the visual N1 component of the visual-only trials as a function of whether they were attended vs. unattended. The visual N1 component was measured over a left and a right parietal–occipital ROI, averaged across four electrode sites on each side (centered approximately over the standard 10–20 sites P3/PO5 on the left and P4/PO6 on the right; see Fig. 2). Repeated-measures ANOVAs were used to test for visual spatial attention effects by testing for significance of the three-way interaction between the factors Attention (attended, unattended), Stimulus location (left, right) and Hemisphere (left, right).

Multisensory spreading of attention

To examine the spreading of attention towards the synchronously occurring auditory input (cf. Busse *et al.*, 2005) we used a series of subtractions to extract the ERP response to the task-irrelevant, centrally presented auditory stimulus (i.e. the spoken letter) as a function of its multisensory context. In the first step, we extracted the auditory ERP when the sound occurred synchronously with an attended vs. an unattended lateral visual letter, collapsed over incongruent and congruent stimulus types, thereby revealing the main effect across time of the multisensory spreading of attention. (See Results and Fig. 3 for details on this extraction process). In the second series of steps, the attentional-spread ERP waveforms were extracted separately for incongruent and congruent multisensory combinations, and contrasted in multiple ways to assess the influence of cross-modal stimulus conflict on the spreading of attention activity.

Results

Behavioral results

For the attended sequential pairs of visual letters A and X that did not match (i.e. the target pairs), subjects responded with a button press after the second item in the pair, reporting whether it was an A or an

X. Correct responses were defined as behavioral responses occurring between 200 and 1000 ms following the onset of the second letter of the sequential pair. The overall accuracy of the subjects collapsed across the multisensory context conditions was relatively high, averaging 91%, which did not differ significantly as a function of condition. The mean RTs collapsed across the multisensory context conditions was 672 ms, and this did vary as a function of condition, as described below.

Grand-average ($n = 16$) mean response times, averaged over the responses to visual letters A and X, were (mean \pm SD) 669 ± 157 ms for the pure (unisensory) visual stimuli, and 648 ± 162 and 700 ± 177 ms for the multisensory congruent and incongruent audiovisual stimuli, respectively. A repeated-measures ANOVA including the factor Condition with three levels (pure, congruent, incongruent) revealed a highly significant effect of multisensory context ($F_{1,45,22.45} = 16.6$, $P < 0.0001$, Greenhouse–Geisser corrected). Specific comparison tests indicated that RTs to the incongruent multisensory targets were significantly slower than to the congruent ones ($F_{1,15} = 55.5$, $P < 0.0001$). Additional specific comparisons also indicated that RTs to pure visual targets were significantly faster than incongruent visual target responses ($F_{1,15} = 7.9$, $P = 0.013$), but were significantly slower than congruent responses ($F_{1,15} = 5.7$, $P = 0.03$), suggesting that there was no general arousal effect of the sound on the visual letters over the multisensory conditions. These behavioral results indicate that, while attending the laterally presented visual letters, subjects appeared to be influenced by the presence of the centrally presented, task-irrelevant incongruent sounds, being distracted (slower) when they were incongruent and facilitated when they were congruent.

EEG results

Evidence of covert visual spatial attention

Indication of selectively focused covert spatial attention on the designated lateralized letter stream was explored by examining visual spatial attention effects on the amplitudes of the parietal–occipital sensory N1 component for the pure visual stimuli, contralateral to the direction of attention. Visual spatial attention effects on the N1 amplitude are shown in Fig. 2, for the left and right parietal–occipital ROIs. Repeated-measures ANOVAs confirmed the presence of robust visual spatial attention effects by a significant three-way interaction between the factors Attention (attended, unattended), Stimulus location

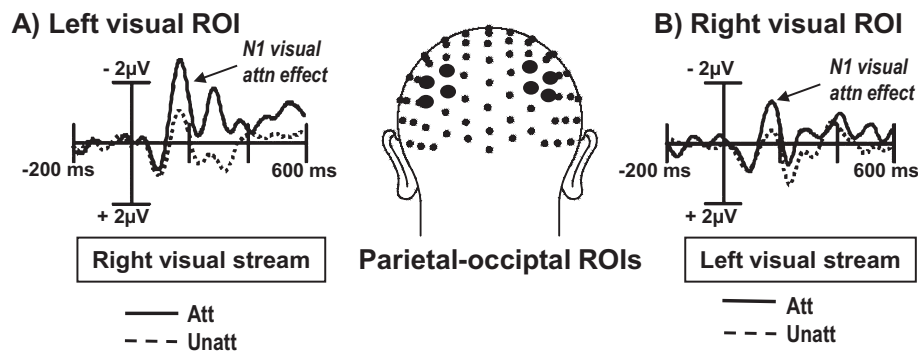


FIG. 2. Evidence for subject's covert visual attention: modulation of early visual sensory activity due to the selectively focused visual spatial attention. (A) Left visual ROI: traces of the attentional modulation of the visual N1 component contralateral to the right side of pure visual stimulation. The traces were revealed by averaging over four electrodes over left parietal–occipital scalp (center panel), separately for when the visual stimuli were attended and unattended. (B) Right visual ROI: traces of the attentional modulation of the visual N1 component contralateral to the left side of pure visual stimulation. The traces were revealed by averaging over four electrodes over right parietal–occipital scalp (center panel), separately for when the visual stimuli were attended and unattended. Note that the enhancement of the N1 sensory response when the visual stimuli were attended.

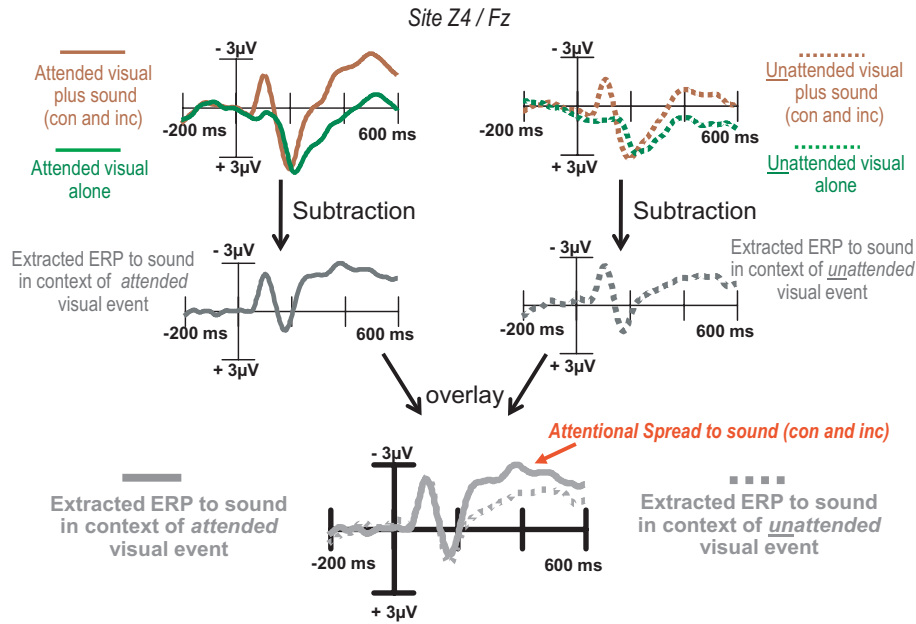


FIG. 3. Schematic illustration of the extraction of auditory ERPs to reveal attentional spreading shown at frontal site Z4 (~Fz). Data were collapsed across congruent (con) and incongruent (inc) sound types. (Left) Attended condition: a mixture of auditory and visual components can be seen in the ERP to the audiovisual letter–sound combinations independent of whether they were coupled to a congruent or incongruent sound; brown trace). The subtraction of the visual-alone ERP (green trace) from the audiovisual ERP yields the ‘extracted’ ERPs to the sound in the context of an attended visual letter collapsed across congruent and incongruent conditions (grey solid line). (Right) Unattended condition: the analogous subtraction is performed on the unattended unisensory visual letters (green dotted trace) and the unattended audiovisual letter–sound combinations (brown dotted line). The grey dotted trace shows the corresponding unattended-condition difference wave of the multisensory minus unisensory-visual ERPs. (Bottom) The extracted difference waves overlaid, revealing the attention-related difference of a task-irrelevant, spatially discordant, spoken letter sound occurring in the context of an attended vs. unattended visual letter stimulus (i.e. attentional spread activity, indicated in orange).

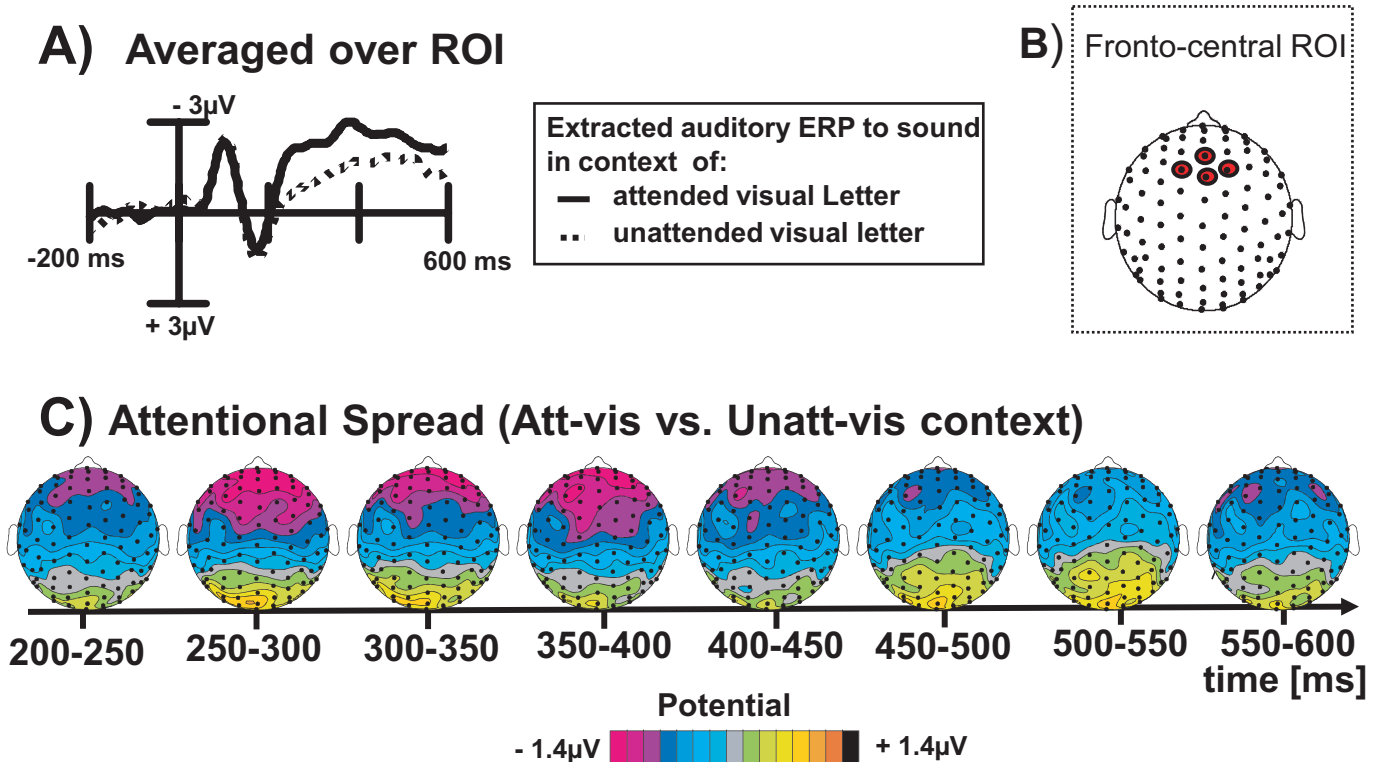


FIG. 4. Spread of attention averaged over congruent and incongruent trials. (A) Traces of extracted auditory ERP averaged over (B) the frontal–central ROI for a centrally presented ignored sound with an attended (line) or unattended (dotted) visual letter from a lateral position. Note that the attended-visual and unattended-visual traces start to diverge at approximately 220 ms, indicating the start of attentional spread. (B) Top view topographical display of the locations of four electrode sites included in the frontal–central ROI (red dots); locations are averaged over 16 subjects. (C) Scalp topographies of the differences of the attended minus unattended traces of A, showing the spread of visual attention, collapsed over congruency conditions.

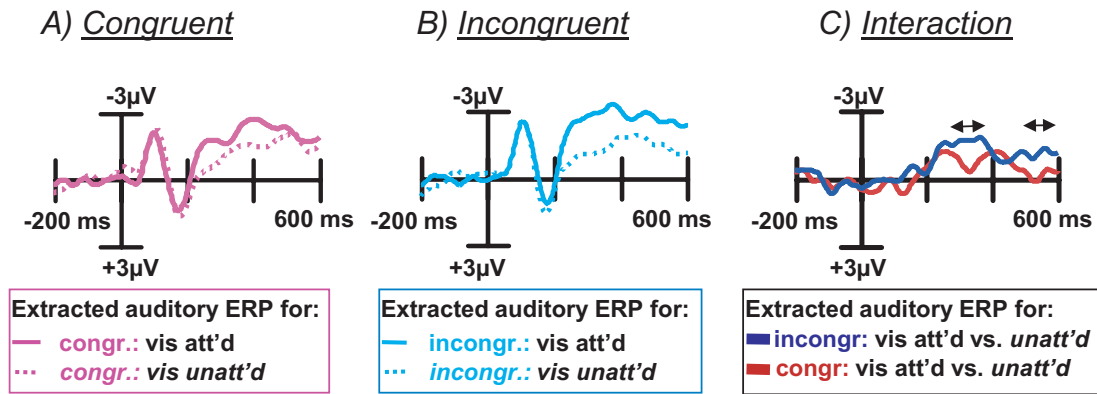


FIG. 5. Traces of the spread of visual attention, shown separately as a function of multisensory incongruity for the frontal–central ROI. (A) Traces of the extracted auditory ERPs for sounds congruent with an attended (solid line) or unattended (dotted line) visual letter, for the frontal–central ROI. (B) Corresponding traces of the extracted auditory ERP for sounds incongruent with an attended (solid line) or unattended (dotted line) visual letter. (C) Traces of extracted auditory ERP for the attentional differences of A and B, thus showing the interaction of attention and congruency. Black arrows mark the intervals of significant interaction.

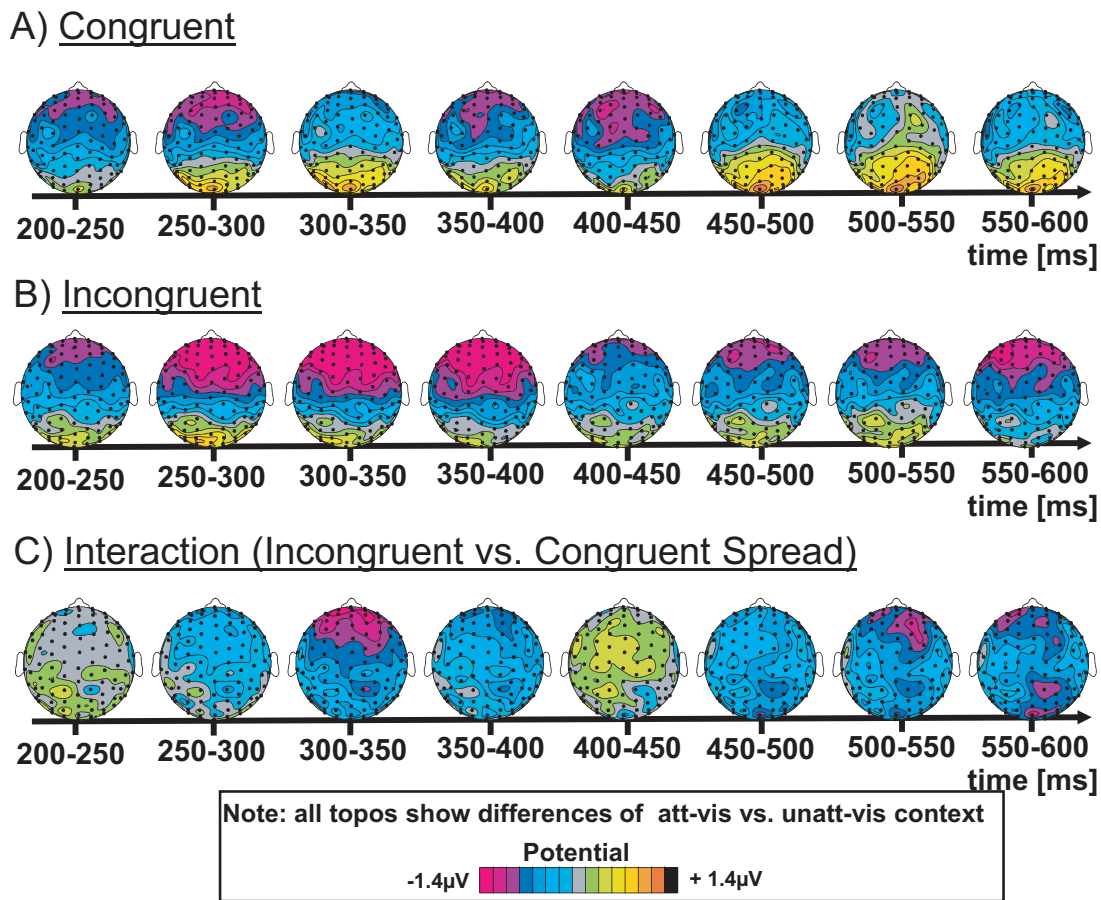


FIG. 6. Scalp topographies of the spread of visual attention, shown separately as a function of multisensory incongruity. (A) Scalp topographies of congruent attentional spread revealed by the difference of congruent sounds with an attended minus unattended visual letter, shown in successive 50-ms bins (see Fig. 4A). (B) Corresponding scalp topographies of incongruent attentional spread revealed by the difference of incongruent sounds with an attended minus unattended visual letter (see Fig. 4B). (C) Scalp topographies of the interaction of attentional spread by congruency, thus indicating the difference given by incongruent (B) minus congruent (A) scalp topographies.

(left, right) and Hemisphere (left, right) ($F_{1,15} = 5.3, P = 0.004$), resulting from the N1 being larger for the visually attended visual letter stimuli. These robust visual attention effects indicate that our participants performed the attentional task as instructed.

Multisensory spreading of attention averaged over congruency conditions

We first extracted auditory ERP activity independent of multisensory context (i.e. averaged across congruent and incongruent trial types) by

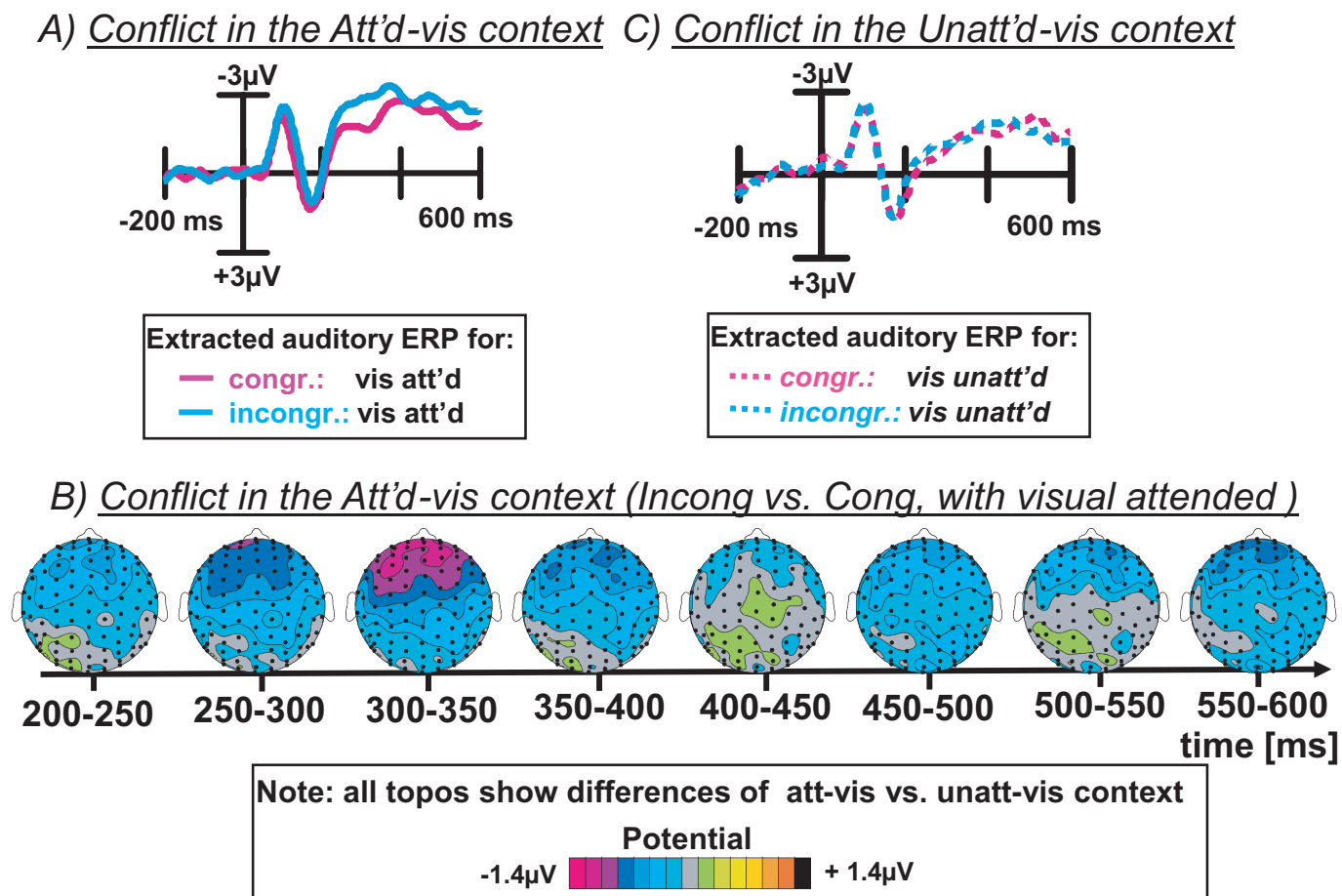


FIG. 7. Traces and corresponding scalp topographies of conflict processing, shown separately as a function of visual attention. (A) Traces of the extracted auditory ERPs for the centrally presented sounds when they either congruent (pink solid line) or incongruent (blue solid line) with a simultaneous, attended, visual letter, shown for the frontal–central ROI. (B) Scalp topographies of the incongruency effects shown in A when the simultaneous visual stimulus was attended, shown in successive 50-ms bins (cf. A). (C) Traces of the extracted auditory ERPs for sounds congruent (pink dotted line) or incongruent (blue dotted line) with a simultaneous, unattended visual letter, shown for the same frontal–central ROI. Note that there was no effect of incongruency when the simultaneous, laterally presented, visual stimulus was unattended.

subtracting activity elicited by attended visual-only (V) trials from activity elicited by the visually attended multisensory (VA) trials (see Fig. 3, left side for the electrode site Z4 [\sim Fz]). The analogous subtraction was also applied to the unattended trial responses (Fig. 3, right side). This subtraction removes the simple visual sensory components and visual attention effects that are in common for the pure visual stimuli and the visual part of the multisensory stimuli. In the second subtraction step, these data were used to compute attended-context minus unattended-context ERP difference waves (Fig. 3, bottom), which isolates activity specific for the ‘spreading-of-attention’ across the component parts of a multisensory object (cf. Busse *et al.*, 2005).

This activity specific for the ‘spreading-of-attention’ over sound and space appeared as a late-onsetting (220 ms), long-lasting, frontally distributed negativity (see traces over the frontal–central ROI in Fig. 4A and scalp-potential distribution maps in Fig. 4C). Within-subjects repeated-measures ANOVAs, including the factor Attention (attended visual vs. unattended visual), revealed that extracted auditory ERPs were significantly more negative over frontal–central scalp sites for spoken letters occurring synchronously with an attended visual event than with an unattended visual event, starting at 220 ms ($F_{1,15} = 17.6$, $P < 0.0008$ between 220 and

240 ms) and ending at 480 ms after stimulus onset, with $F_{1,15}$ values ranging between 17.6 and 6.3, and corresponding P -values between 0.0008 and 0.025).

Multisensory attentional spread as a function of congruency

Our main goal in the present study was to examine interaction effects of higher-level representation-related stimulus conflict on the multisensory spreading of attention. In the paragraphs below, we first discuss attentional-spread ERP activity across time separately for the incongruent and congruent multisensory conditions. Then we discuss the interaction effects between the two multisensory context conditions.

Specific analyses for the congruent stimuli (Figs 5A and 6A) showed significant spreading of attention activity over the frontal–central ROI (as revealed by a significant increased negativity for the attended, compared to the unattended, condition), but occurring in two, somewhat separated, time ranges. The first time range started at 220 ms and persisted until 300 ms. Tested within 20-ms windows of averaged data over the frontal–central ROI, main effects of multisensory attentional context were found with $F_{1,15}$ values ranging between 18.7 and 6.6 (corresponding P -values between 0.0006 and 0.022). The second time period started at 340 ms and persisted until 440 ms (main

effects of multisensory attentional context, $F_{1,15}$ values ranging between 7.6 and 4.7, corresponding P -values between 0.015 and 0.05).

For the incongruent stimuli (Figs 5B and 6B), within-subject repeated-measures ANOVAs, including the factor Attention (attended-visual, unattended-visual) and applied to the frontal–central ROI, revealed a significant difference in the frontal negativity between the attended-visual and unattended-visual conditions (i.e. the attentional spreading activity) starting at approximately 220 ms and persisting continuously until 500 ms (main effects of multisensory attentional context, $F_{1,15}$ values ranging between 37.4 and 5.9, corresponding P -values between 0.0001 and 0.03 for the 20-ms windows). No significant effects were found between 500 and 540 ms, but significant differences reoccurred between 540 and 600 ms ($F_{1,15}$ values ranging between 8.5 and 11.2, corresponding P -values between 0.004 and 0.015).

In order to reveal interaction effects of attention and conflict, we computed the ‘spreading of attention’ ERP activity (attended vs. unattended extracted auditory ERPs) separately for the incongruent and the congruent trial types (double difference waves in Figs 4C and 5C), and compared these. This contrast indicated that the spreading-of-attention negativity was larger for incongruent than for congruent multisensory stimuli over frontal–central scalp sites (Figs 5C and 6C). The corresponding within-subject repeated-measures ANOVA, applied to data from the frontal–central ROI, revealed a significant interaction between Attention (attended-visual, unattended-visual) and Congruency started at 300 ms and persisted until 360 ms ($F_{1,15}$ values ranging between 7.0 and 6.0, corresponding P -values between 0.02 and 0.03 across 20-ms windows), with a trend towards significance between 360 and 380 ms ($F_{1,15} = 4.0$, $P = 0.065$). After this initial phase, the interaction effect reappeared for a short time period between 540 and 560 ms ($F_{1,15} = 5.0$, $P = 0.04$). Thus, attention drawn towards the auditory part of the multisensory object appeared to be larger during processing of conflicting, nonmatching letter–sound combinations than for matching ones.

The effect of incongruency as a function of multisensory attentional context

As another way of examining the interaction between attentional context and conflict-related activity, we examined the effects of the incongruency of the task-irrelevant auditory stimuli as a function of the multisensory attentional context (see Fig. 7A and C). As can be seen from the figure, a conflict effect (incongruent minus congruent stimuli) was only present in the attended-visual condition, not in the unattended-visual one. This observation was confirmed by specific comparisons showing significant conflict activity over the frontal–central ROI when the visual stimulus was attended (as revealed by a significant increased negativity for the incongruent, compared to the congruent, condition), which was manifested in two separated time ranges [220–380 and 520–600 ms; (P -values between 0.05 and 0.001, analyzed in 20-ms bins)]. In contrast, there were no significant differences for the analogous comparisons for the visual-unattended conditions. Thus, conflict processing for the task-irrelevant auditory stimuli occurred when the simultaneous visual stimulus was attended, and not when it was unattended.

Discussion

In the present study, we investigated the amplitude and temporal characteristics of the ‘spreading-of-attention’ across a multisensory

audiovisual object when the auditory and visual stimulus components not only were spatially discordant but also conflicted at a semantic or representational level. To examine the spreading of attention toward the synchronously occurring auditory input (cf. Busse *et al.*, 2005), we extracted auditory ERP waves by subtracting visual-alone activity from multisensory activity (VA), separately for the attended-visual and unattended-visual conditions, which could then be compared. First, collapsed across visual–auditory congruency conditions, these activation patterns replicated our previously reported results from Busse *et al.* (2005). More specifically, a spread of attention was observed from the visual toward the auditory part of a multisensory stimulus that was comprised of spatially discordant, but simple, auditory and visual stimulus components. Secondly, the spreading of attention in the present study occurred for both the congruent and incongruent conditions, starting at the same point in time (220 ms). Third, interactions of this spreading of attention activity and multisensory conflict at a higher representational level began at a later point in time, starting at approximately 300 ms, manifested as increased activity for the incongruent than the congruent stimulus combinations. Moreover, specific comparisons analyzing this interaction indicated that the visual–auditory conflict effects were only present when the laterally-presented visual component was attended. These findings suggest that as attention spread across both modality and space to the centrally presented, task-irrelevant, auditory stimuli, these auditory stimuli tended to attract even greater attention when they were incongruent with the task-relevant attended visual stimulus. Furthermore, these incongruency effects interacting with attention spreading occurred some time after attention had begun to spread from the visual to the auditory stimulus part, suggesting that the higher-level conflict was detected after the accrual of some multisensory interaction processes at a semantic or representational level.

Multisensory spread of attentional collapsed over congruency relationships

Collapsed over incongruent and congruent multisensory trial types, the extracted auditory ERPs showed attentional spreading activity starting at approximately 220 ms, reflected as an enhanced, sustained, negative wave over frontal–central scalp for the task-irrelevant central spoken sounds when they occurred simultaneously with an attended vs. an unattended lateral visual-letter stimulus. This effect was very similar to the effect reported over comparable scalp sites by Busse *et al.* (2005), in which the task-relevant lateral visual stimuli were simple geometric figures and the task-irrelevant centrally presented auditory stimuli simple tones (i.e. neither meaningfully related stimuli nor incongruent). In addition, as was the case in the Busse *et al.* (2005) study, no effects of multisensory attentional context (i.e. whether the synchronous visual stimulus was attended or not) were found on the auditory response earlier than the frontal negativity starting after 200 ms. This is in contrast with studies in which the auditory stimulus occurs in the spatial locus of attention (either visually directed or auditorily directed spatial attention), which typically results in an enhanced auditory N1 sensory component at 100 ms, relative to when spatial attention is directed elsewhere (Woldorff & Hillyard, 1991; Talsma & Woldorff, 2005). Thus, this multisensory late-negativity effect can be interpreted as attention having to spread first from the visual to the spatially separated but synchronous auditory stimulus component, which apparently takes approximately 200 ms or so, before any attention-related effects can be expressed in the auditory extracted ERPs (Busse *et al.*, 2005). Our current data, collapsed across congruency, therefore replicated these earlier results.

Multisensory spread of attention as a function of visual–auditory incongruency

Analysis of the spreading-of-attention activity separately for congruent and incongruent stimulus combinations indicated that it started at the same time (approximately 220 ms) in the two conditions, thus being independent of whether the stimulus components were congruent or not. Later in time, however, the attentional-spreading activity was larger for the incongruent stimulus combination than for the congruent, a difference that began at approximately 300 ms and was significant during two later windows (300–380 and 540–560 ms).

In regards to the observed amplitude differences in the spreading-of-attention activity between incongruent and congruent combinations, we had hypothesized two alternative possible outcomes. One possibility was that attention would spread more to the task-irrelevant auditory stimulation for the congruent than for the incongruent trials, the rationale being that once the incongruency was detected the brain would respond by sending inhibitory signals to suppress the processing of the conflicting auditory stimulus. The alternative hypothesis was that the incongruent auditory stimulation would serve as a greater distractor, capturing attentional resources and thus leading to a higher amplitude response in the extracted auditory ERPs. Our results showing increased spreading of attention for the incongruent stimulus trials support the alternative hypothesis. In addition, these incongruency differences were only found when the visual component was attended, not when it was unattended. Thus, our results would be consistent with various previous unimodal visual ERP studies using higher level, top-down, stimuli (Flanker paradigm: Wendt *et al.*, 2007; Bartholow *et al.*, 2005; Heil *et al.*, 2000; Stroop paradigm: Liotti *et al.*, 2000; Appelbaum *et al.*, 2009; West & Alain, 1999; Badzakova-Trajkov *et al.*, 2008) that have shown increased processing for the incongruent (conflicting) vs. congruent stimulus combinations. We interpret the present results as reflecting an increased tendency for the task-irrelevant incongruent auditory stimulus to attract a greater level of attention.

In regards to the temporal onset of the spreading–conflict interaction, we suggested two possible outcomes. One was that there could be a relatively immediate influence of conflict on spreading, in that conflict effects and attentional spreading effects show similar onsets, at least as reported across studies [e.g. Wendt *et al.*, 2007 (conflict) vs. Busse *et al.*, 2005 (spread)]. Another possibility was that the influence of conflict on attentional spreading might be somewhat delayed, under the view that the simultaneous auditory stimuli might have to be drawn at least partially into the penumbra of the visual attention, and/or to reach a higher enough processing level after receiving some attention, before its conflicting nature would be detected and differentially responded to. Our results provide evidence for the delay hypothesis (conflict interaction effect being delayed, here until 300 ms), suggesting that conflict was detected after some spreading of attention to the task-irrelevant incongruent auditory stimulus had occurred.

Relationship to supramodally focused spatial attention paradigms

In most previous EEG and MEG studies investigating conflicting vs. matching auditory–visual stimuli (e.g. face–voice, Stekelenburg & Vroomen, 2007; picture–sound, Fiebelkorn *et al.*, 2010; Yuval-Greenberg & Deouell, 2007; letter–sound, Raij *et al.*, 2000; Herdman *et al.*, 2006), the stimuli were generally presented from the same spatial location, typically centrally (but see, with fMRI, Fairhall & Macaluso, 2009). In such studies, spatial attention was therefore already in place for the location of both the visual and auditory stimuli

(i.e. supramodally) and did not need to spread spatially to encompass the simultaneously occurring auditory stimulus. When the auditory stimuli are presented at the same location as the attended visual stimulus, however, it cannot be distinguished whether any observed multisensory conflict effects depend on attention, at least on spatial attention, or could occur also without being in the spatial attentional focus (i.e. there is no unattended spatial condition to compare to).

Note that in the present study the task-irrelevant auditory stimulus was always the same, and always occurred centrally, with the multisensory attentional manipulation being whether the simultaneous lateral visual stimulus was in a spatially attended location or a spatially unattended one. The contrast between these thus enables an isolation of the attentional spreading effect. Moreover, our data indicate that there is multisensory conflict interaction only when the simultaneous task-relevant visual component is presented in the spatially attended location, with no effect of incongruency when it occurs in a spatially unattended location.

Stimulus-related vs. representation-related multisensory attention effects

In a recent study by Fiebelkorn *et al.* (2010) investigating multisensory conflict processing of over-learned image–sound combinations it was proposed that the multisensory spreading attention could either be mainly stimulus-driven or representation-driven. In that study, all stimuli were presented centrally at a spatially attended location, but with only the images being task-relevant. The images were of three different objects types (dogs, cars and guitars), and in separate runs one of these types was designated the target object type, with the task to focus their object attention on those and perform a one-back working memory task. On some trials, a task-irrelevant sound was presented that could be either congruent with the image (a barking sound with a dog) or incongruent (a car sound with a dog), and on some trials the image was presented alone. Using a series of ERP subtractions, the authors aimed to separate stimulus-driven attentional spreading from representation-driven attentional spreading, the former being a bottom-up process dependent on physical stimulus properties and the latter dependent on the task-relevance of the stimuli (target or nontarget). For the stimulus-driven spreading in that study, an onset timing of approximately 240 ms over frontal–central electrodes sites was found, but the amplitude of this activity in the 200–300 ms time range was independent of the identity (incongruent or congruent) of the auditory and visual components. Thus, in this time range, our results agree with the Fiebelkorn *et al.* (2010) ones, showing no effect of incongruency on this initial, bottom-up, attentional-spreading activity, derived from the simultaneous occurrence of the visual and auditory components.

On the other hand, we observed greater attentional-spreading activity later on, beginning at approximately 300 ms, for the incongruent stimulus combinations. While the study of Fiebelkorn *et al.* (2010) focused on the initial time period for this effect, prior to 300 ms, showing no influence of incongruency there, the attentional spreading effect in their data lasted for an extended period of time, as it did in our study. Of potential relevance is that if one looks carefully at their data traces it would appear that, later on, as in our data, there was substantially greater attentional spreading activity for the incongruent condition (compare traces for congruent ‘audiovisual (AV-V)’ nontargets in their fig 4a and incongruent ‘audiovisual (AV-V)’ nontargets in their fig 4b: Fiebelkorn *et al.*, 2010, page 114), with this enhanced activity for the incongruent condition running from

approximately 300 ms until the end of their figure traces at 500 ms. The authors do not report statistical results from this later latency range, and it may well have only been a trend, but such a difference would appear to be quite consistent with our results.

In terms of their suggested model of distinguishing between stimulus-driven and representation-driven attentional spreading, we would interpret our current results in the following way: this attentional spreading effect begins in a bottom-up, stimulus-driven way, with initially no difference as a function of congruency. However, after the attentional spread has proceeded for some period of time, the auditory and visual component processing reaches a representational level of analysis, at which time their conflicting nature is detected, triggering this later conflict-related difference. Thus, in this view, the stimulus-driven attentional spread is indeed initially fully bottom-up, unrelated to representation and just resulting from the visual and auditory component co-occurrence, but an interaction can occur when the processing reaches a representational level. We note that this influence of representation would appear to be different from the representation-related effect that Fiebelkorn *et al.* (2010) were focusing on, in which the representation-driven attentional spreading is due to the top-down, representation-based attention to a specific category of object type (e.g. dogs vs. cars), with this task-related top-down manipulation inducing its own multisensory attentional spreading activity. As in our study we focused on a manipulation of visual spatial attention rather than on a manipulation of attending to one type of object vs. another (i.e. in our case, this might have been comparing attending to A vs. attending to X), our data cannot comment on this other aspect of representation-driven, fully top-down task-related effects examined by Fiebelkorn *et al.* (2010).

Semantically-related N400 effects and the influence of spatial attention

In the present study the comparison of the response to an incongruent vs. a congruent central auditory letter stimulus that was paired with a visual stimulus in a spatially attended location gave a robust incongruency effect, but no such effect of incongruency was seen when the visual stimulus was in the unattended location (see design in Fig. 1 and traces in Fig. 7). An ERP activation to which our results may be related is the N400 wave, which is a late negative-polarity component peaking at approximately 400 ms post-stimulus that is sensitive to semantic incongruity (Kutas & Hillyard, 1980). Importantly, the N400 is also an indicator of the semantic incongruity of two sequentially presented words or pictures, exhibiting a greater amplitude in response to unrelated words or pictures than to related ones (reviewed in Van Petten & Luka, 2006). A previous study by McCarthy & Nobre (1993) investigated the interplay of spatial attention and semantic processing by presenting two visual streams of words to different spatial location on the screen, one of which was spatially attended and the other unattended. The typical N400 negativity peaking at 400 ms, revealed by semantically unrelated vs. related words, could only be found in the spatially attended visual stream and not in the spatially unattended one. Analogously, in the present study we found that multisensory 'semantic-object' letter processing also depended on attention, in our case due to the multisensory attentional context (i.e. it only occurred when the simultaneous, spatially discordant, visual stimulus component was attended). On the other hand, the typical N400 distribution tends to be more posteriorly distributed (i.e. central-parietal, Van Petten & Luka, 2006) than the more anterior (frontal-central) distribution we find for the negative-polarity enhancement here. Nevertheless, it is possible

that the attention-dependent, multisensory, letter-incongruency effect reported here may bear some relationship to other previously reported, attention-dependent, negative-polarity activations that are sensitive to semantic incongruency more generally.

Summary and conclusion

In summary, we investigated 'spreading of attention' across a multisensory audiovisual object when the auditory and visual stimulus components were spatially discordant and semantically conflicting. Attentional spreading started at approximately 220 ms over frontal-central sites, for both congruent and incongruent trial types, as reflected by a slow frontal-central negativity, replicating the results that were reported in Busse *et al.* (2005) for simple visual and auditory stimuli that were neither meaningfully related nor unrelated. In addition, this frontal-central negativity was specifically larger at later time points (300–360 and 540–560 ms) for incongruent than for congruent stimulus components. We conclude that the increased frontal-central negativity for the incongruent condition, which occurred some time after the onset of attentional spreading, indicates an increased attentional capture for this condition that occurs after attention has spread from the task-relevant visual to the task-irrelevant but conflicting auditory stimulus part, presumably as a result of some degree of multisensory higher-level processing interactions. Future studies, using methods with higher spatial resolution (e.g. fMRI), will be required to investigate whether this effect reflects increased activity only in auditory sensory cortex (cf. Busse *et al.*, 2005) for the incongruent auditory stimulus, or whether it also includes additional activity from other brain areas in frontal cortex, such as the anterior cingulate cortex, that have been implicated as being involved in conflict detection and resolution.

Acknowledgements

We thank Joseph Harris for helpful editorial input on the manuscript. This research was supported by NSF grant 0524031 & NIH grant ROI-NS051048 to M.G.W.

Abbreviations

ERP, event-related potential; ROI, region of interest; RT, reaction time; SOA, stimulus-onset-asynchrony.

References

- Appelbaum, L.G., Meyerhoff, K.L. & Woldorff, M.G. (2009) Priming and backward influences in the human brain: processing interactions during the Stroop interference effect. *Cereb. Cortex*, **19**, 2508–2521.
- Badzakova-Trajkov, G., Barnett, K.J., Waldie, K.E. & Kirk, I.J. (2008) An ERP investigation of the Stroop task: the role of the cingulate in attentional allocation and conflict resolution. *Brain Res.*, **1253**, 139–148.
- Bartholow, B.D., Pearson, M.A., Dickter, C.L., Sher, K.J., Fabiani, M. & Gratton, G. (2005) Strategic control and medial frontal negativity: beyond errors and response conflict. *Psychophysiology*, **42**, 33–42.
- Busse, L., Roberts, K.C., Crist, R.E., Weissman, D.H. & Woldorff, M.G. (2005) The spread of attention across modalities and space in a multisensory object. *Proc. Natl. Acad. Sci. U S A*, **102**, 18751–18756.
- Eriksen, B.A. & Eriksen, C.W. (1974) Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.*, **16**, 143–149.
- Fairhall, S.L. & Macaluso, E. (2009) Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *Eur. J. Neurosci.*, **29**, 1247–1257.
- Fan, J., Flombaum, J.I., McCandliss, B.D., Thomas, K.M. & Posner, M.I. (2003) Cognitive and Brain Consequences of Conflict. *Neuroimage*, **18**, 42–57.

- Fan, J., Kolster, R., Ghajar, J., Suh, M., Knight, R.T., Sarkar, R. & McCandliss, B.D. (2007) Response anticipation and response conflict: an event-related potential and functional magnetic resonance imaging study. *J. Neurosci.*, **27**, 2272–2282.
- Fiebelkorn, I.C., Foxe, J.J. & Molholm, S. (2010) Dual mechanisms for the cross-sensory spread of attention: how much do learned associations matter? *Cereb. Cortex*, **20**, 109–120.
- Hanslmayr, S., Pastötter, B., Bäuml, K.H., Gruber, S., Wimber, M. & Klimesch, W. (2008) The electrophysiological dynamics of interference during the Stroop task. *J. Cogn. Neurosci.*, **20**, 215–225.
- Heil, M., Osman, A., Wiegmann, J., Rolke, B. & Henninghausen, E. (2000) N200 in the Eriksen-Task: inhibitory executive process? *J. Psychophysiol.*, **14**, 218–225.
- Herdman, A.T., Fujioka, T., Chau, W., Ross, B., Pantev, C. & Picton, T.W. (2006) Cortical oscillations related to processing congruent and incongruent grapheme-phoneme pairs. *Neurosci. Lett.*, **399**, 61–66.
- Kutas, M. & Hillyard, S.A. (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, **207**, 203–205.
- Liotti, M., Woldorff, M.G., Perez, R. & Mayberg, H.S. (2000) An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia*, **38**, 701–711.
- McCarthy, G. & Nobre, A.C. (1993) Modulation of semantic processing by spatial selective attention. *Electroencephalogr. Clin. Neurophysiol.*, **88**, 210–219.
- Raij, T., Uutela, K. & Hari, R. (2000) Audiovisual integration of letters in the human brain. *Neuron*, **28**, 617–625.
- Stekelenburg, J.J. & Vroomen, J. (2007) Neural correlates of multisensory integration of ecologically valid audiovisual events. *J. Cogn. Neurosci.*, **19**, 1964–1973.
- Stroop, J.R. (1935) Studies of interference in serial verbal reactions. *J. Exp. Psychol.*, **18**, 643–662.
- Talsma, D. & Woldorff, M.G. (2005) Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *J. Cogn. Neurosci.*, **17**, 1098–1114.
- Van Atteveldt, N.M., Formisano, E., Goebel, R. & Blomert, L. (2007) Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage*, **36**, 1345–1360.
- Van Petten, C. & Luka, B.J. (2006) Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain Lang.*, **97**, 279–293.
- Van Veen, V. & Carter, C.S. (2002) The timing of action-monitoring processes in the anterior cingulate cortex. *J. Cogn. Neurosci.*, **14**, 593–602.
- Wendt, M., Heldmann, M., Münte, T.F. & Kluwe, R.H. (2007) Disentangling sequential effects of stimulus- and response-related conflict and stimulus-response repetition using brain potentials. *J. Cogn. Neurosci.*, **19**, 1104–1112.
- West, R. & Alain, C. (1999) Event-related neural activity associated with the Stroop task. *Brain Res. Cogn. Brain Res.*, **8**, 157–164.
- Woldorff, M.G. & Hillyard, S.A. (1991) Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalogr. Clin. Neurophysiol.*, **79**, 170–191.
- Yeung, N., Botvinick, M.M. & Cohen, J.D. (2004) The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychol. Rev.*, **111**, 931–959.
- Yuval-Greenberg, S. & Deouell, L.Y. (2007) What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. *J. Neurosci.*, **27**, 1090–1096.