



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

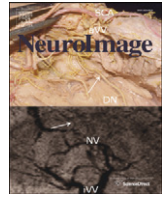
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

NeuroImage

journal homepage: [www.elsevier.com/locate/ynimg](http://www.elsevier.com/locate/ynimg)

# Early multisensory interactions affect the competition among multiple visual objects

Erik Van der Burg<sup>a,\*</sup>, Durk Talsma<sup>b</sup>, Christian N.L. Olivers<sup>a</sup>, Clayton Hickey<sup>a</sup>, Jan Theeuwes<sup>a</sup>

<sup>a</sup> Vrije Universiteit Amsterdam, The Netherlands

<sup>b</sup> University of Ghent, Belgium

## ARTICLE INFO

### Article history:

Received 1 June 2010

Revised 22 December 2010

Accepted 24 December 2010

Available online 31 December 2010

### Keywords:

Audiovisual integration

Event-related potential

Visual search

Attention

## ABSTRACT

In dynamic cluttered environments, audition and vision may benefit from each other in determining what deserves further attention and what does not. We investigated the underlying neural mechanisms responsible for attentional guidance by audiovisual stimuli in such an environment. Event-related potentials (ERPs) were measured during visual search through dynamic displays consisting of line elements that randomly changed orientation. Search accuracy improved when a target orientation change was synchronized with an auditory signal as compared to when the auditory signal was absent or synchronized with a distractor orientation change. The ERP data show that behavioral benefits were related to an early multisensory interaction over left parieto-occipital cortex (50–60 ms post-stimulus onset), which was followed by an early positive modulation (80–100 ms) over occipital and temporal areas contralateral to the audiovisual event, an enhanced N2pc (210–250 ms), and a contralateral negative slow wave (CNSW). The early multisensory interaction was correlated with behavioral search benefits, indicating that participants with a strong multisensory interaction benefited the most from the synchronized auditory signal. We suggest that an auditory signal enhances the neural response to a synchronized visual event, which increases the chances of selection in a multiple object environment.

© 2010 Elsevier Inc. All rights reserved.

## Introduction

Many studies have reported that information from different sensory modalities interacts (Alais and Burr, 2004; Calvert et al., 2000; Giard and Peronn  t, 1999; Hershenson, 1962; Macaluso et al., 2000; McGurk and MacDonald, 1976; Molholm et al., 2002; Schroeder and Foxe, 2005; Shipley, 1964). For instance, a single visual event is perceived as being brighter when accompanied by an auditory signal than when presented in isolation (Stein et al., 1996). These and other results have provided evidence for the notion that multisensory integration enhances signal clarity and/or reduces stimulus ambiguity (see e.g. Chen and Yeh, 2009; Olivers and Van der Burg, 2008; Vroomen and De Gelder, 2000). One drawback of the majority of studies to date, however, is that they examine interactions among single events at a time (i.e. a single visual event in combination with a single auditory event), thus leaving out the question how multisensory interactions can aid in resolving the competition between multiple stimuli (Spence, 2007).

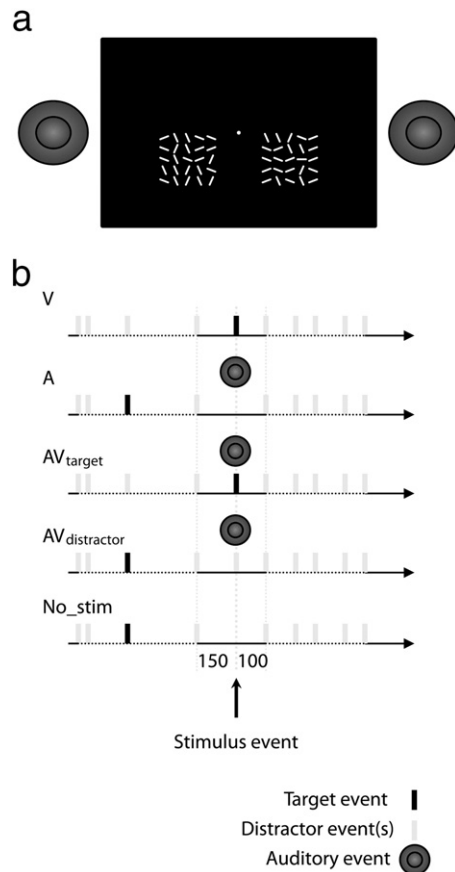
One exception is provided by a recent study of ours (Van der Burg et al., 2008b) in which participants searched for a horizontal or vertical line segment presented among many distractor lines of various orientations. Throughout each experimental trial random subsets of items changed color. Search through these displays is difficult but

improves dramatically when a spatially uninformative tone is concurrently presented with the color change of the target line. This “pip and pop” effect, as we have dubbed it, indicates that a synchronized auditory event can affect the competition among multiple visual items. Follow-up studies demonstrated that the pip and pop effect is not due to increases in alertness or top-down temporal cueing (Van der Burg et al., 2008a,b). We have proposed that the auditory signal enhances the neural response of the synchronized visual event at an early sensory level of processing (Stein et al., 1996), though in our prior study we could not provide direct evidence for this hypothesis.

Here we report an electrophysiological study in humans that supports the idea that early multisensory integration underlies the pip and pop effect, bolstering the claim that rapid and automatic audiovisual integration leads to the subjective experience of the synchronized visual event popping out of the background. Participants were asked to search for a horizontal or vertical target line among irrelevant diagonal lines. At pseudo-random intervals subsets of lines changed orientation. This set-up allowed us to independently control the moment of display onset and the moment of target appearance within that display. The target was placed at a lateral location in the lower visual field so as to take advantage of the known lateralization of brain potentials related to the deployment of attention. The target was absent at the beginning of the trial, and at a specific moment one distractor line changed into the target line (i.e. a horizontal or vertical line segment). Participants were required to identify the target orientation by pressing one of two keys. Our behavioral measure was identification accuracy. Fig. 1a presents an example search display (see on-line Mov. 1 for a video clip of a trial).

\* Corresponding author. Cognitive Psychology, Vrije Universiteit Amsterdam, Van der Boerhorststraat 1, 1081BT Amsterdam, The Netherlands.

E-mail address: [e.van.der.burg@psy.vu.nl](mailto:e.van.der.burg@psy.vu.nl) (E. Van der Burg).



**Fig. 1.** (a) Illustration of the search display used in the present study. During each trial, a random number (1, 4 or 7) of distractor lines changed orientation, and one of them changed into the target line (i.e. a horizontal or vertical line segment). Participants were instructed to make an unspeeded response to the orientation of this target line. Two loudspeakers were located on the left and the right of the monitor. (b) Illustration of the different stimulus conditions. The event related potential (ERP) was time-locked to the stimulus event. The light gray bars indicate the moments at which a number (1, 4 or 7) of distractor elements changed orientation. The black bar indicates the moment that a single distractor line changed into the target line (i.e. a horizontal or vertical line segment). The event of interest, the synchronized event, was always preceded by a 150 ms interval and followed by a 100 ms interval. Furthermore, all other intervals varied randomly between 50 and 250 ms. See online Mov. 1 for an example display used in the present study (AV<sub>target</sub> condition).

The current experiment was designed in such a way that the relation between multisensory integration and attention could be studied from two different perspectives. First, ERPs elicited by auditory (A) and visual (V) signals presented in isolation could be summated and compared to those ERPs elicited by audiovisual (AV) stimuli, in which the sound coincided with the visual target (i.e. investigating the difference between AV and [A + V] ERPs), without specifically considering the lateral position of the visual stimuli. Earlier ERP work has shown that this additive model can reveal early latency (~40 ms) multisensory processes (Giard and Peronn  t, 1999; Mishra et al., 2007; Molholm et al., 2002; Talsma et al., 2007), reflecting changes early in the visual processing sequence (Eckert et al., 2008; Martuzzi et al., 2007; Zangenehpour and Zatorre, 2010). The present study is thus designed to determine whether such early modulations underlie the pip and pop effect.

To assess the possibility that a sound automatically interacts with a coinciding visual stimulus (regardless of its relevance (Van der Burg et al., 2008a,b)) we not only included a condition in which the visual part of the audiovisual stimulus was the target (i.e. an AV<sub>target</sub>), but also a condition in which the visual part of the audiovisual stimulus was an irrelevant distractor (AV<sub>distractor</sub>). Here the tone was synchronized (and ERPs time-locked) to a visual distractor stimulus. If the pip and pop effect occurs in an

automatic fashion, then early modulations are expected here as well as in the AV<sub>target</sub> condition. Note that the different conditions were presented in blocks to ensure that participants ignored the auditory event in the AV<sub>distractor</sub> condition. This way, interactions could consequently be attributed to automatic integration of multisensory information. Fig. 1b illustrates the temporal dynamics of the trial types employed in our study.

Second, in addition to analyzing early multisensory modulations, the presentation of visual events at lateral locations in the visual field allowed for the analysis of lateralized ERPs associated with sensory enhancement, the deployment of attention, and activation in of visual short term memory (VSTM). To index changes in these cognitive processes we looked at a series of well-documented ERP components: the lateral P1 (90–120 ms post-stimulus), the N2pc (175–300 ms post-stimulus), the contralateral negative slow wave (CNSW; 300+ ms post-stimulus).

Multimodal research with the lateralized P1 has associated increases in the amplitude of this component to modulation of sensory strength and priority (McDonald et al., 2005; St  rmer et al., 2009). In McDonald et al. (2005), for example, the perceived order of lateralized visual events was affected by the location of a preceding auditory signal. Visual events that were presented at the same location as auditory events were perceived as occurring earlier in time, and this behavioral effect was associated with increase in the amplitude of the lateral P1. These results lead us to expect a corresponding increase in lateral P1 in the present study if the auditory signal in fact modulates the strength of a synchronized visual event (though it is important to note that, in contrast to McDonald et al. (2005), in the present study the sound itself was not lateralized).

The N2pc was originally linked to visuo-spatial selective attention in a series of visual search studies (Luck and Hillyard, 1994; Luck et al., 2000). The component is evident as an increase in negative ERP amplitude at posterior electrodes located over cortical brain areas contralateral to an attended visual object. In the present study, we expect to observe an N2pc if the synchronized visual event captures attention. Moreover, if the early audiovisual integration and subsequent change in visual processing occurs automatically, similar P1 and N2pc modulations should be observed in the AV<sub>distractor</sub> and AV<sub>target</sub> conditions.

The CNSW has been associated with visual short term memory (VSTM; Klaver et al., 1999). This lateralized component is present during memorization and reflected as a posterior slow negative ERP wave developed over the hemisphere contralateral to the memorized item. In our experimental task, VSTM was presumably active in AV<sub>target</sub> trials, as participants were required to eventually respond based on target characteristics, but we did not expect the recruitment of VSTM resources in the AV<sub>distractor</sub> condition. With this in mind we expected to find a larger CNSW in the AV<sub>target</sub> condition.

Finally, we expect to observe a greater P3 component in the AV<sub>target</sub> condition compared to the other conditions (Sutton et al., 1965). We distinguish the P3 from co-occurring CNSW in that the P3 is a central component, while the CNSW has a lateral topography. The P3 has been associated with awareness of stimuli, and/or updating of working memory (Nieuwenhuis et al., 2005; Sutton et al., 1965).

To foreshadow, our results show that spatially uninformative auditory signals start to affect the competition among multiple visual objects at an early sensory stage (50–60 ms after stimulus onset), driving an early positive modulation contralateral to the visual event (~80 ms). These effects are followed by enhanced N2pc, contralateral negative slow wave (CNSW; Klaver et al., 1999), and P3 components, reflecting the orienting process of attention, visual short term memory, and active cognitive processing.

## Materials and methods

### Participants

Fourteen volunteers participated in the experiment (7 females, mean age 23.8 years; range 18–33 years). Participants were either



paid € 10 an hour or received course credits. All participants gave written informed consent for their participation. Experimental procedures adhered to the declaration of Helsinki.

### Design and stimuli

Two grids filled with randomly oriented line elements were presented in the lower left and right visual fields (see Fig. 1a). A trial lasted approximately 4 seconds, and during this period a number of line elements changed orientation. At the start of the trial, all line elements were diagonally oriented. After a randomly determined duration, one of the changing lines ended up in either horizontal or vertical orientation. This element was designated as the target stimulus. At the end of the trial, participants were required to report the orientation of the target (horizontal or vertical) by making an unspeeded response. On some trials, a brief tone could coincide with the target change.

More specifically, the present study contained five different stimulus conditions (V, A, AV<sub>target</sub>, AV<sub>distractor</sub>, and No-stim; see Fig. 1b). In the visual only (V) condition, sounds were absent and ERPs were time-locked to the target orientation change. In the auditory only (A) condition, a sound was present but did not coincide with a visual event. In the AV<sub>target</sub> condition sounds were presented simultaneously with the target orientation change. In the AV<sub>distractor</sub> condition, sounds were presented simultaneous with an orientation change of one of the distractor elements. In the No-stim condition, ERPs were locked to a time-point at which there was neither a visual change nor a sound. Although trials in this condition did consist of a stimulus sequence similar to that in the other conditions, in this condition the ERPs were locked to a point in time where nothing happened. Since target events were still present (randomly before or after the time-locking non-event), participants were still engaged in performing the task. Therefore, these No-stim trials could be used to estimate brain activity related to generic cognitive processes not specific to the multisensory integration processes of interest and remove the corresponding ERP activity from the trials of interest. In addition, these No-stim trials served to estimate the residual activity elicited by adjacent distractor changes in the sequence of visual display changes (see Talsma and Woldorff, 2005b, for a similar methodology).

The auditory stimulus was a 500 Hz tone (44.1 kHz sample rate, 16 bit, mono) with duration of 60 ms (including a 5 ms fade-in and 5 ms fade-out to avoid audible transients) presented via two speakers placed such that the sounds appeared to originate from the center of the screen. The visual search display consisted of 50 white ( $76.7 \text{ cd m}^{-2}$ ) line segments (length  $0.60^\circ$  visual angle) on a black ( $<0.05 \text{ cd m}^{-2}$ ) background. The orientation of each line element was pseudo-random. Half of the lines were placed on an invisible  $5 \times 5$  grid ( $3.40^\circ \times 3.40^\circ$ ) in the lower left visual field ( $5.32^\circ$  to the left and  $2.13^\circ$  below fixation), and the other half of the lines were placed on an identical grid in the lower right visual field ( $5.32^\circ$  to the right and  $2.13^\circ$  below fixation). Participants were required to maintain fixation at the center of the screen, which was marked by a white ( $76.7 \text{ cd m}^{-2}$ ) dot. The orientation of each line deviated randomly by either plus or minus  $22.5^\circ$  from horizontal or vertical, except for the target, which changed from a diagonal to a horizontal or vertical line during the course of a trial. During each trial, twenty orientation changes occurred. The length of the interval between changes varied randomly from 50 to 250 ms, with the exception that the event of interest (to which ERPs were time-locked) was always preceded by a 150 ms interval and followed by a 100 ms interval to promote unambiguous audiovisual binding (Van der Burg et al., 2010b, 2008b). This event of interest (see stimulus event in Fig. 1b) always occurred randomly between the 7th and 14th orientation change. In the No-stim, AV<sub>distractor</sub>, or A condition we time-locked the ERPs to events other than the visual target event. In that case, the visual target event occurred on the fourth or fifth change before or after the event of interest. Target stimuli could never occur on the outer ring of the  $5 \times 5$  grid.

During each orientation change, one, four, or seven line segments changed orientation (90 degrees, either left- or rightwards). The orientation change that revealed the target (i.e. when one of the elements changed to a horizontal or vertical orientation) was unique in that during that moment only one element (the target itself) changed orientation, and it changed by less than 90 degrees (i.e. sufficient to align it with horizontal or vertical).

### Procedure

Each trial began with a fixation dot that was presented for 1000 ms at the center of the screen. This was followed by the presentation of the search array and the sequence of 20 orientation changes. At the end of the trial, participants were instructed to make an unspeeded response to indicate whether the target orientation was vertical or horizontal. Participants were instructed to respond with high accuracy. Target orientation (horizontal or vertical) as well as its location (left or right visual field) were balanced and randomly mixed within blocks. In the AV<sub>distractor</sub> condition, the location (left or right visual field) of the synchronized distractor was also balanced and randomly mixed within blocks. Participants received five practice blocks, and 85 experimental blocks (5 conditions  $\times$  17 blocks) of 16 trials each, leading to 272 trials per relevant cell (collapsed across vertical and horizontal target orientations). After each block, participants received accuracy feedback.

### Apparatus

The experiment was run in a dimly lit, sound-attenuated, electrically shielded cabin. Participants were seated approximately 80 cm from the monitor and speakers (which were placed on the left and right side of the monitor). EEG was recorded with 128 active sintered Ag-AgCl electrodes (BioSemi, Amsterdam, The Netherlands) positioned radially equidistant from the vertex across the scalp (BioSemi 128 channel ABCD layout; see <http://www.biosemi.com> for details). Additional electrodes were placed at the left and right mastoids, approximately 1 cm above and below the orbital ridge of each eye, and at the outer canthi of the eyes. Eye movements were also monitored using a video monitoring system. Data was digitized at 512 Hz and referenced during recording to an active common mode electrode and rereferenced offline to the average of the signals recorded at the two mastoids.

Horizontal eye-movements were detected by rereferencing the electrodes at the left and right canthi against each other, thus reconstructing the horizontal electro-oculogram (EOG). Likewise, vertical eye movements and blinks were detected by referencing the electrodes above each eye against the electrode directly below it. This was done for each eye separately, thus reconstructing the vertical EOGs.

### Analysis

The behavioral data was subjected to a repeated measures within-subjects univariate ANOVA with stimulus condition (V, A, AV<sub>target</sub>, AV<sub>distractor</sub>, No-stim) as within-subject factor and alpha set at .05. The reported *P* values have been Huynh-Feldt corrected for sphericity violations.

Raw EEG signals were digitally filtered using a 0.05 Hz, 4096-point Gaussian high-pass filter, and an 18 Hz, 201-point Blackman windowed finite impulse low-pass filter. Epochs containing eye-blinks (vEOG:  $150 \mu\text{V}/500 \text{ ms}$ ) and eye-movements (hEOG:  $75 \mu\text{V}/20 \text{ ms}$ ; vEOG:  $75 \mu\text{V}/20 \text{ ms}$ ) were discarded using an automated procedure (Talsma and Woldorff, 2005a). Furthermore, the signal-to-noise ratio in ERP data was optimized by using an auto-adaptive procedure (Talsma, 2008). ERP data were analyzed for correct response trials only, unless otherwise stated, resulting in the exclusion of approximately 25% of trials. Of all correct trials, an additional 15% were rejected due to artifacts. This left on average at least 150 trials for each cell. In each

condition, ERPs were obtained by time-locking to the event of interest in that condition and averaging all qualifying EEG epochs. These averages contained a 200 ms pre-stimulus baseline period and extended until 1000 ms after the event of interest. ERPs were baseline corrected using the 200 ms pre-stimulus period. To control for differential overlap, and generic cognitive processes, the time-locked averages elicited by the no-stim trials were first subtracted from each of the original (A, V, and AV<sub>target</sub> and AV<sub>distractor</sub>) ERPs (Talsma and Woldorff, 2005b).

The resulting ERP averages were subjected to several analyses. First, multisensory integration effects were estimated by summing the A and V (A + V) ERPs and comparing this summated ERP to the ERP elicited by the AV<sub>target</sub> and AV<sub>distractor</sub> trials (Giard and Peronn  t, 1999; Stein and Meredith, 1993). While we expect an early (< 100 ms) multisensory interaction, we did not have a clear prediction regarding the location of the early effect. Therefore, for correct trials only, we initially conducted an explorative analysis for the early multisensory integration effect, testing each sample (~2 ms) and electrode for an effect of the within subjects factor Stimulus Type (i.e. AV<sub>target</sub>, AV<sub>distractor</sub>, and (A + V)); analysis was limited to the first 200 ms following stimulus onset; cf. Giard and Peronn  t, 1999). These results were thresholded such that effects were only considered to be significant when a  $P$  value < 0.05 was found on at least two sequential samples at two or more neighboring channels. These criteria were chosen because they represented the minimum temporal and spatial thresholds necessary to remove some spurious results from our analyses (i.e. choosing more restrictive criteria did not change our results appreciably). After this overall analysis, three post-hoc tests were conducted, in which each of the individual levels of the factor Stimulus Type was tested against one of the other levels. Results from these tests were considered significant when they adhered to the same criteria as applied to the overall test (i.e.  $P$  < 0.05 on at least two significant samples and at two adjacent channels), with the added restriction that that any time point/channel combination that had yielded no significant result on the overall test was discarded. Fig. 4a illustrates how the early multisensory interaction evolves over time. Of main interest is the earliest multisensory interaction effect. This interaction was reliable at around ~50 ms (post stimulus) and most pronounced over the left hemisphere (see Fig. 5; 50–60 ms time window).

This early effect was further examined in detail by obtaining a mean voltage across the time window that garnered significant results in the exploratory analysis (i.e., 50–60 ms; see Fig. 5). These mean voltages were used as dependent measures in an ANOVA with a within-subjects factor for Stimulus Type (A + V, AV<sub>target</sub> vs. AV<sub>distractor</sub>). For correct trials only, separate ANOVAs were used to examine the presence of an early multisensory interaction in the AV<sub>target</sub> condition (i.e. AV<sub>target</sub> vs. (A + V)) and AV<sub>distractor</sub> condition (i.e. AV<sub>distractor</sub> vs. (A + V)). Note that the AV<sub>distractor</sub> condition was compared against the same A + V conditions as the AV<sub>target</sub> condition, even though in the AV<sub>distractor</sub> condition the ERPs were time-locked to a visual distractor change rather than a visual target change. This is because the raw exogenous visual response elicited by the distractor and target orientation changes should not differ. To test whether any behavioral benefits were indeed related to the early multisensory interaction, we conducted a Post-hoc analysis for correct and incorrect response trials in the AV<sub>target</sub> condition, under the assumption that on incorrect trials multisensory integration would fail more often.

The later P3 effect was tested by computing mean amplitudes, averaged across three medial parietal electrodes (A19, A20, A21; approximately equivalent to PZ), at consecutive 10 ms time windows. These values were subjected to an ANOVA containing the within subjects factors Stimulus Type (AV<sub>target</sub> vs. A + V).

Whereas the analysis of the early multisensory integration effect was somewhat explorative, the analysis of subsequent lateralized effects and the central P3 component were hypothesis driven. Therefore, we did not conduct any exploratory analyses involving all electrodes, but only

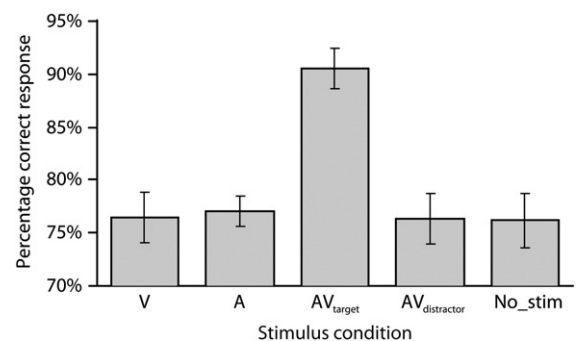
analysed these effects using electrode locations where these effects are expected to be maximal. The presence of lateralized effects was determined by creating three separate sets of ERPs for the AV<sub>target</sub>, AV<sub>distractor</sub>, and V<sub>only</sub> trials, depending on whether the visual stimulus occurred in the left or right visual field. The resulting waveforms were tested for the presence of an early P1, a late N2pc, and a late memorization effect by obtaining from left and right hemifield trials (in the AV<sub>target</sub>, AV<sub>distractor</sub>, and V<sub>only</sub> conditions) the mean voltages of consecutive 10 ms time windows. These mean voltages were averaged across three posterior left (approximately equivalent to P3, PO5, PO7) and three posterior right (approximately equivalent to P4, PO6, PO8) hemisphere electrodes (as identified in Figs. 8 and 9), and were used as dependent measures in an ANOVA containing the within-subjects factors Stimulus Location (target left vs. target right), Stimulus Type (AV<sub>target</sub>, AV<sub>distractor</sub>, vs. V<sub>only</sub>), and Hemisphere (left vs. right).

BESA (BESA International, Germany) was used to create topographies of the ERP effects. The lateral ERP scalp topography maps (early contralateral positivity and N2pc) are based on an contralateral-minus-ipsilateral difference mirrored across the vertical meridian of the electrode array with electrodes on the midline artificially set to zero (see also Hickey et al., 2009). For topographic mapping the ERP data was re-referenced to the average voltage observed at 642 positions across a spherical-spline-interpolated representation of the data (i.e. in BESA nomenclature, the data was transformed such that it was “reference-free”). The average reference value was subtracted from the contra-minus-ipsilateral difference waves to make the reference-free maps show zero on the midline.

## Results

### Behavioral data

Target detection performance differed significantly for the different stimulus conditions,  $F_{4,56} = 31.1$ ,  $P < 0.0001$ . As is clear from Fig. 2, the auditory signal led to strong benefits for the detection of the synchronized visual target in the AV<sub>target</sub> condition, resulting in significantly improved accuracy (90.5%) compared to all other conditions (average 75.8%; all individual comparisons,  $t_{13} > 5.9$ , all  $P$ s < 0.0001). As in previous studies (Ngo and Spence, 2010; Van der Burg et al., 2010b; 2008a,b, 2009), spatially uninformative auditory events affected the competition between multiple visual items. The current results also demonstrate that the pip and pop phenomenon is not restricted to target color changes, indicating that the multisensory integration process underlying the effect is not feature specific. Further analysis yielded no reliable difference in performance when we excluded the AV<sub>target</sub> condition,  $F_{3,39} < 1$ ,  $P > 0.8$ . Note that a synchronized distractor orientation change (AV<sub>distractor</sub> condition) did not lead to behavioral costs, as might be expected if the AV<sub>distractor</sub> drew attention to its location. However, it is important to note that the



**Fig. 2.** Mean percentage of trials in which participants reported the target orientation correctly as a function of stimulus condition. Data were collapsed across target orientation and target location. The error bars represent the overall standard errors of individual mean correct responses.

interval between the visual target and the synchronized distractor event was quite long (200–2500 ms; i.e. the minimum and maximum possible time interval between the distractor and target change), raising the possibility that attention may have been initially deployed to the distractor and later reoriented to the target, with no behavioral consequence (e.g. Theeuwes, 2000). Furthermore, although there were no costs for the  $AV_{distractor}$  on average, a more detailed analysis revealed a reliable correlation across participants between the benefits in the  $AV_{target}$  condition (relative to the V condition) and the costs in the  $AV_{distractor}$  condition (also relative to the V condition), Pearson's  $r = .627$ ,  $P < 0.02$ , indicating that those who benefited the most from audiovisual targets suffered the most from audiovisual distractors.

#### ERP data

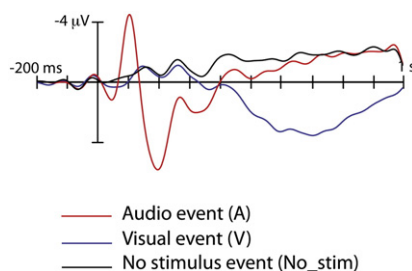
Fig. 3 presents the ERPs elicited when the auditory event was presented alone (A), when the visual event was presented alone (V), and when neither event occurred (No-stim).

The auditory (A) ERP has a standard morphology with clearly identifiable P50 and N100 components. In contrast, the visual (V) ERP contained no strong exogenous components such as the P1 and N1. This is consistent with studies showing attenuation of these early components when dynamic visual displays are employed (Leblanc et al., 2008; Martens et al., 2007).

The ERP results for the multisensory events revealed a sequence of relevant modulations, beginning with an early component over left parieto-occipital cortex (50–60 ms post-stimulus) for both targets and distractors. This component was only apparent for correct responses, consistent with the idea that it reflects the integration of modality-specific information that underlies the pip and pop effect. This was followed by a series of lateralized ERP effects: an early (80–120 ms) positivity contralateral to the AV-target and AV-distractor events, suggesting a sensory enhancement of the multisensory event compared to the unisensory visual stimuli, and an N2pc, reflecting stimulus selection. When the multisensory event defined a target, the N2pc was followed by a large CNSW, demonstrating the activation of VSTM, and a larger P3 component, possibly reflecting enhanced awareness of the audiovisual target. These results demonstrate that the auditory stimulus, though spatially uninformative, rapidly integrated with the concurrently presented visual stimulus. This led to benefits in subsequent perceptual, attentive, and mnemonic operations.

#### Early multisensory integration

An initial exploratory analysis yielded reliable multisensory interactions between ~50 and 180 ms (see Fig. 4 for a statistical overview). Fig. 4a details the statistical significance of  $AV_{target} \times AV_{distractor} \times (A + V)$  amplitude for all electrodes within the 0–200 ms interval (for correct trials only). Fig. 4b and c represents the multisensory interaction for



**Fig. 3.** Grand average ERPs elicited by the auditory event alone (A), visual event alone (V), or when there was no auditory or visual event (No\_stim). Data are collapsed across the same three posterior left-hemisphere electrodes as in Fig. 5 (see 50–60 ms interval). Note that negative is plotted upward and that stimulus onset occurred at 0 ms, which is indexed by the y-axis.

$AV_{target}$  and  $AV_{distractor}$  trials, respectively. Note that in this latter analysis we only considered effects that were reliable in the preceding omnibus analysis (as depicted in Fig. 4a).

As is clear from Fig. 4a, the earliest reliable interaction among the  $AV_{target}$ ,  $AV_{distractor}$  and sum of the unisensory conditions ( $A + V$ ) was apparent from about ~50 ms. This was most pronounced over the left hemisphere (see Fig. 5; 50–60 ms time window). This early interaction was examined in detail by an ANOVA with Stimulus type ( $A + V$ ,  $AV_{target}$  vs.  $AV_{distractor}$ ) as within subjects factor. This ANOVA yielded a reliable main effect of Stimulus Type,  $F_{2,26} = 5.31$ ;  $P < 0.05$ , (50–60 ms). This main effect was further examined by two separate ANOVAs to investigate early multisensory interactions in the  $AV_{target}$  and  $AV_{distractor}$  condition.

*Early multisensory integration in the  $AV_{target}$  condition (50–60 ms).* Fig. 6 presents the ERPs for correct and incorrect trials to multisensory AV-target stimuli compared to the sum of the ERPs to unisensory A and V stimuli ( $A + V$ ) and the corresponding scalp topographies (Fig. 5a) for those trials on which observers responded correctly.

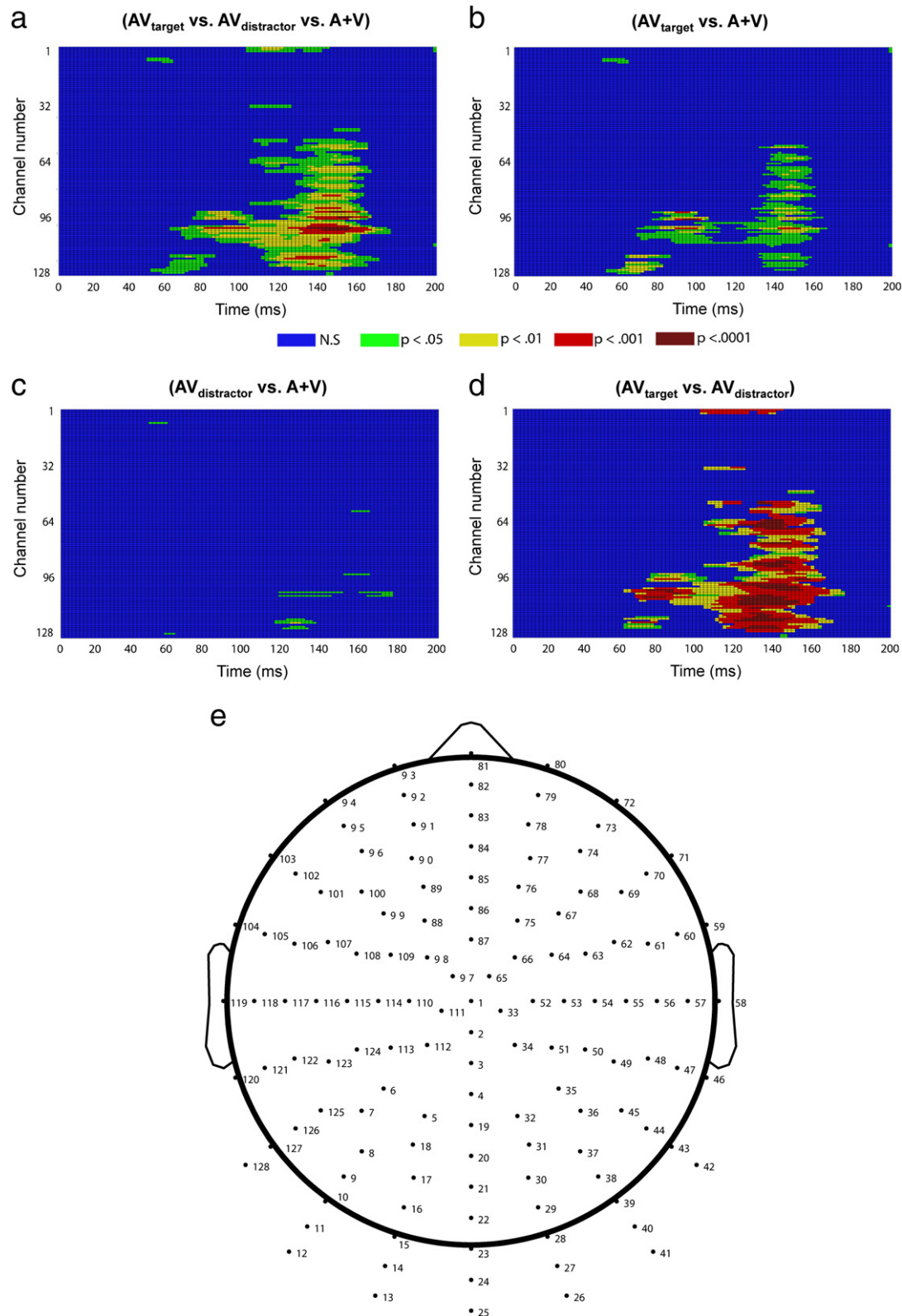
In the  $AV_{target}$  condition (for correct trials only), the main effect of Stimulus type ( $AV_{target}$  vs.  $A + V$ ) was reliable,  $F_{1,13} = 6.81$ ,  $P < 0.05$ . This early multisensory interaction is lateralized and most pronounced over the left parieto-occipital electrodes (see the 50–60 ms interval in Fig. 5a). The latency of the early multisensory interaction is consistent with earlier work suggesting that an auditory signal can affect the processing of a synchronized visual event early in the cortical processing stream (Giard and Peronn  t, 1999; Molholm et al., 2002; Talsma et al., 2007; Talsma and Woldorff, 2005b).

The early multisensory integration effect in the  $AV_{target}$  condition was further examined for correct and incorrect trials to investigate whether any behavioral benefits were related to the multisensory integration effect. We expected that in incorrect trials multisensory integration would fail more often than in correct trials (see Fig. 6). This was statistically confirmed by a Stimulus Type ( $A + V$  vs.  $AV_{target}$ )  $\times$  Response accuracy (correct response vs. incorrect response) interaction,  $F_{1,13} = 7.10$ ;  $P < 0.05$ . Furthermore, the main effect of Stimulus Type ( $A + V$  vs.  $AV_{target}$ ) was not reliable when participants responded incorrectly,  $F_{1,13} = 3.68$ ,  $P > 0.3$ . These results suggest that the early component plays a crucial role in causing efficient detection of the  $AV_{target}$  (i.e. the pip and pop effect). A more detailed analysis additionally revealed a reliable correlation across participants between the benefits in the  $AV_{target}$  condition (relative to the V condition) and the magnitude of the early multisensory interaction ( $AV_{target}$  vs.  $A + V$ ), Pearson's  $r = -.615$ ,  $P = 0.019$ , indicating that those with a strong multisensory interaction benefited the most from the audiovisual targets (see Fig. 7). This correlation bolsters the claim that the early multisensory interaction reflected in the ERP plays a crucial role in creating the pip and pop behavioral effect.

*Early multisensory integration in the  $AV_{distractor}$  condition (50–60 ms).* If the early process of integration is truly automatic, we should expect this component to emerge when the AV event identifies a distractor event ( $AV_{distractor}$  condition). Fig. 6 presents the ERPs for correct trials to multisensory  $AV_{distractor}$  stimuli compared to the sum of the correct ERPs to unisensory A and V stimuli ( $A + V$ ) and the corresponding scalp topographies (Fig. 5b). The ANOVA yielded a reliable Stimulus Type ( $A + V$  vs.  $AV_{distractor}$ ) effect,  $F_{1,13} = 5.71$ ,  $P \leq 0.05$ . This confirms our hypothesis that multisensory integration occurs automatically, because in the experimental blocks where the AV event identified a distractor, it never identified a target. There was therefore no strategic reason for the participants to integrate the audio and visual modalities.

Note that in the  $AV_{distractor}$  condition we included only those trials on which the target event was presented after the AV event, which constituted half of all trials. This ensured that participants were still maintaining the visual search task as in the  $AV_{target}$  condition and there was no interference from target-related processing. Furthermore, in

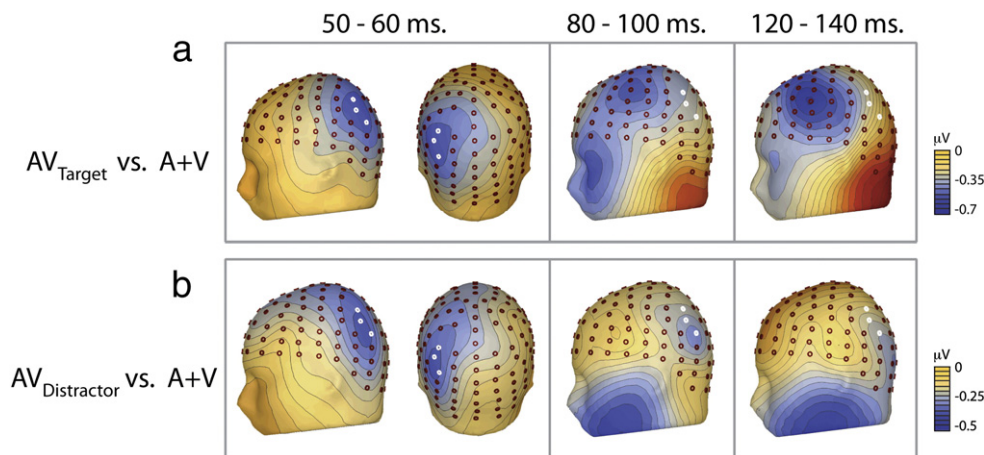




**Fig. 4.** Summary of statistics. Shown here is the statistical significance of the multisensory interactions over all electrodes (biosemi 128 channels) between 0 and 200 ms post-stimulus for all subjects (correct trials only). (a) Statistical significance for the overall interaction ( $AV_{\text{target}} \times AV_{\text{distractor}} \times (A+V)$  at each latency). (b) Statistical significance for multisensory interaction in the  $AV_{\text{target}}$  condition ( $AV_{\text{target}} \times (A+V)$  at each latency). (c) Statistical significance for multisensory interaction in the  $AV_{\text{distractor}}$  condition ( $AV_{\text{distractor}} \times (A+V)$  at each latency). (d) Statistical difference between the  $AV_{\text{target}}$  and the  $AV_{\text{distractor}}$  condition. (e) Electrode location with corresponding channel numbers. Note that channel number corresponds to the biosemi layout (1–32 = A1–A32; 33–64 = B1–B32; 65–96 = C1–C32; 97–128 = D1–D32; see <http://www.biosemi.com> for more details).

these trials participants responded with ~70% accuracy, yielding very few incorrect trials. Therefore, we did not conduct an ANOVA with Stimulus type and Response Accuracy (correct vs. incorrect response).

Nevertheless, the reliable main effect of Stimulus Type for correct trials in the  $AV_{\text{distractor}}$  condition confirms our hypothesis that multisensory integration occurs automatically, since the auditory signal never

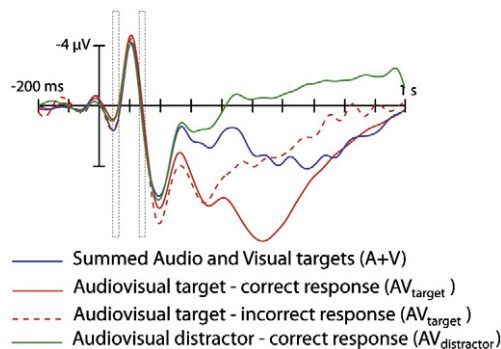


**Fig. 5.** Multisensory interactions. (a) The scalp topographies represent the difference wave created by subtracting the  $AV_{\text{target}} - (A + V)$  ERP. (b) The scalp topographies represent the difference wave created by subtracting the  $AV_{\text{distractor}} - (A + V)$  ERP.

indicated the target event. Moreover, in direct comparison of the multisensory components elicited in the  $AV_{\text{distractor}}$  and  $AV_{\text{target}}$  conditions (for correct trials) there was no evidence of a reliable difference in amplitude,  $F_{1,13} < 1$ ;  $P > 0.4$  (see Figs. 4d and 6).

Some studies criticized the use of an additive model (i.e. comparing the multisensory condition with the sum of the unisensory conditions) to find support for early multisensory interactions, since these early interactions are often due to some “common” activity, like anticipatory brain potentials or motor responses instead (see e.g. Cappe et al., 2010; Teder-Salejari et al., 2002). However, in the present study, responses to the target were unspeeded, and in the  $AV_{\text{distractor}}$  condition we observed a similar multisensory interaction while participants never responded to the synchronized distractor event. Furthermore, anticipatory activation was minimized, since (a) the synchronized event (i.e. the target or distractor in combination with the tone) appeared after a randomized and rather lengthily interval and (b) No-stim trials (see Materials and methods) were first subtracted from each of the original (A, V, and  $AV_{\text{target}}$  and  $AV_{\text{distractor}}$ ) ERPs to control for any residual activity elicited by adjacent distractor changes in the sequence of visual display changes (see also Talsma and Woldorff, 2005b). Therefore, we think that our early interaction reflects a multimodal interaction instead of some common activation.

**Subsequent multisensory integration effects (~80–160 ms).** Fig. 4 illustrates how the early multisensory integration effect evolves over time. Interestingly, the early multisensory integration effect is present for the  $AV_{\text{target}}$  and  $AV_{\text{distractor}}$  condition (50–60 ms), but only present in the



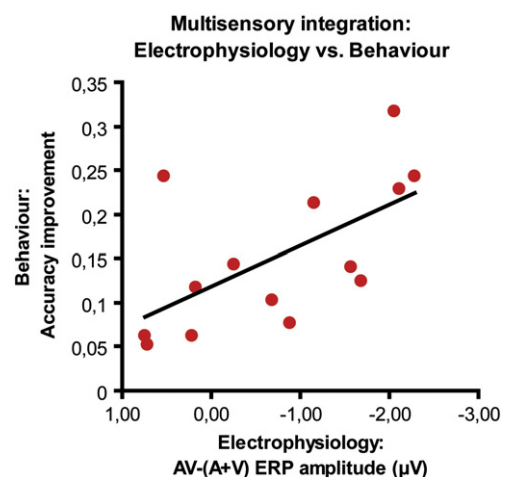
**Fig. 6.** Multisensory interactions in the  $AV_{\text{target}}$  and  $AV_{\text{distractor}}$  condition. Grand average ERPs elicited by the  $AV_{\text{target}}$  stimuli (correct and incorrect trials),  $AV_{\text{distractor}}$  stimuli (correct trials), compared to the sum of the ERPs elicited by A and V stimuli. Data are collapsed across three posterior left-hemisphere electrodes, as indicated by white circles in the scalp topography plot shown in Fig. 5 (see 50–60 ms interval). Note that negative is plotted upward and that stimulus onset occurred at 0 ms, which is indexed by the y-axis.

$AV_{\text{target}}$  condition later in time. This difference suggests that the auditory predictability (i.e. tone synchronized with the target vs. tone synchronized with a distractor) might have had an influence on the prolongation of the multisensory interaction effect. So, attending to the auditory signal in the  $AV_{\text{target}}$  condition does lead to a longer lasting multisensory interaction effect in the  $AV_{\text{target}}$  condition compared to the  $AV_{\text{distractor}}$  condition (in which participants ignored the tone). This sustained multisensory effect was most pronounced over the frontal (80–100 ms) and fronto-central electrodes (140–160 ms; see Fig. 5).

#### Lateralized effects

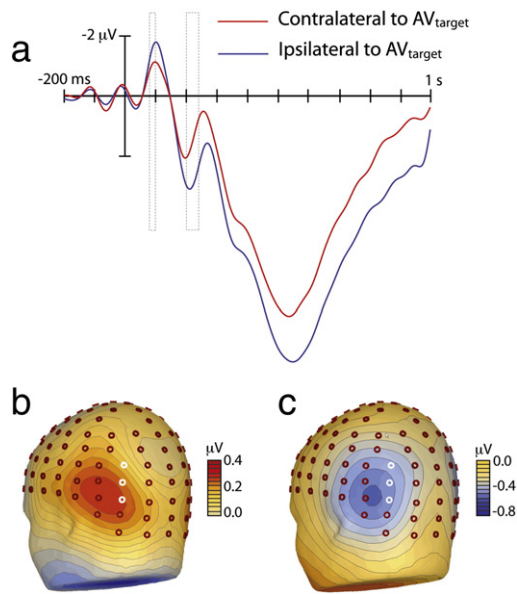
The ANOVA yielded an early and two late reliable Stimulus Location (target left vs. target right)  $\times$  Stimulus Type ( $AV_{\text{target}}$ ,  $AV_{\text{distractor}}$  vs.  $V_{\text{only}}$ )  $\times$  Hemisphere (left vs. right) interactions, at two consecutive 10 ms time windows (80–100 ms;  $F_{1,13} = 4.95$ – $6.25$ ,  $P < 0.05$ ), four consecutive 10 ms time windows (210–250 ms;  $F_{1,13} = 5.69$ – $9.81$ ,  $P < 0.01$ ), and many consecutive 10 ms time windows (330–1000 ms;  $F_{1,13} = 3.44$ – $16.28$ ,  $P < 0.05$ ), respectively. These three interactions were further examined, and discussed as early contralateral positivity, attentional selection as reflected in the N2pc, and lateralized VSTM effects.

**Early contralateral positivity (80–100 ms).** Fig. 8 presents grand averaged target ERPs and scalp topographies for correct trials



**Fig. 7.** Relation between behavioral and electrophysiological correlates of multisensory integration. Accuracy improvement in the  $AV_{\text{target}}$  condition (compared to the V only condition) correlates with the amplitude of the early latency (50–60 ms) ERP effect reflecting the difference between the  $AV_{\text{target}}$  and summated the unisensory conditions (A + V).



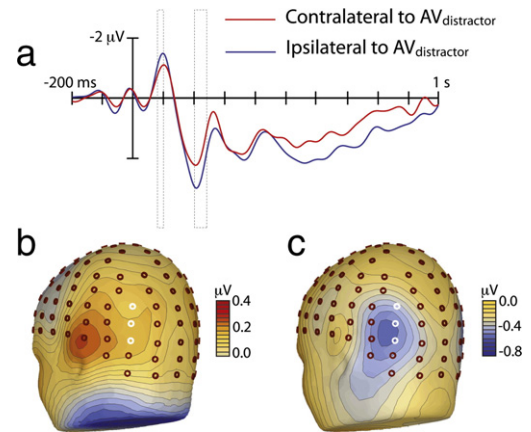


**Fig. 8.** Grand-averaged ERPs and scalp topographies in the AV<sub>target</sub> condition. (a) The ERP waveforms shown here are collapsed across three posterior left and three posterior right hemisphere electrodes, locations of the original electrodes are highlighted by black circles in the scalp topography plots. The grand averaged ERPs represent the target ERPs collapsed over left and right target and left and right hemispheres so as to yield waveforms recorded contralateral and ipsilateral to the side of the synchronized visual event. The scalp topographies represent the difference between contralateral and ipsilateral activity based on the average ERP effect observed across the early interval identified in panel a. (b) Topographical map of the early lateralized effect. (c) Topographical map of the late lateralized effect based on the average ERP effect observed across the late interval identified in panel a. For display purposes, the lateral ERP scalp topography maps (N2pc and early laterality) are based on an contralateral-minus-ipsilateral difference mirrored across the vertical meridian of the electrode array with electrodes on the midline artificially set to zero (see also Hickey et al., 2009).

collapsed across left and right hemifield AV<sub>target</sub> conditions and left and right hemispheres so as to yield waveforms recorded contralateral and ipsilateral to the side of the synchronized visual event.

The early lateralized interaction was further examined by separate ANOVAs for each Stimulus Type (AV<sub>target</sub>, AV<sub>distractor</sub>, vs. V<sub>only</sub>). In the AV<sub>target</sub> condition, the Stimulus Location  $\times$  Hemisphere interaction was reliable,  $F_{1,13} = 23.76$ – $24.00$ ,  $P_s < 0.0001$ , at two consecutive time windows (80–100 ms). This component follows the early multisensory effect depicted in Fig. 5a (see 50–60 ms interval), and is distributed over temporal–occipital areas contralateral to the visual target and has positive polarity (Fig. 8b). In the AV<sub>distractor</sub> condition, we observed a similar reliable, Stimulus Location  $\times$  Hemisphere interaction (see Fig. 9a and b),  $F_{1,13} = 15.48$ – $19.25$ ,  $P_s < 0.001$ , at two consecutive time windows (80–100 ms). In the V<sub>only</sub> condition (see Fig. 10), the Stimulus Location  $\times$  Hemisphere interaction failed to reach significance during this 80–100 ms time window,  $F_{1,13} = 2.01$ – $3.58$ ,  $P_s > 0.0526$ . Further analysis revealed no difference between the AV<sub>target</sub> and AV<sub>distractor</sub> conditions,  $F_{1,13} = 1.13$ – $1.69$ ,  $P_s > 0.216$ .

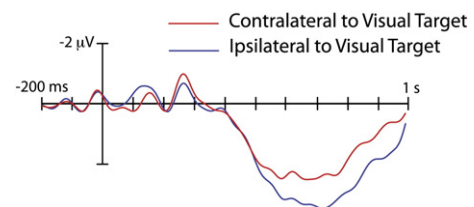
The existing ERP literature typically reports that the auditory N1 component peaks at approximately the same latency as the visual P1 component (i.e. 100 ms). In the present study, we indeed observed a rather strong ERP response to the salient auditory stimulus (see Figs. 8a and 9a) that was characterized by a fronto-central scalp distribution, but was nevertheless still observable over the more posterior recording sites. In contrast, we observed a relatively weak ERP response to the individual visual stimuli (probably due to the dynamic nature of the displays). This situation leads us to believe that the apparent change in N1 activity may in fact reflect a modulation of activity corresponding to the visual P1. The unimodal ERP literature has shown that the P1 is larger in amplitude in response to high contrast stimuli or stimuli with unique perceptual features (Luck and Hillyard, 1994) or when a stimulus is attended (Luck



**Fig. 9.** Grand-averaged ERPs and scalp topographies in the AV<sub>distractor</sub> condition. (a) The ERP waveforms shown here are collapsed across three posterior left and three posterior right hemisphere electrodes, locations of the original electrodes are highlighted by black circles in the scalp topography plots. The grand averaged ERPs represent the distractor ERPs collapsed over left and right distractor location and left and right hemispheres so as to yield waveforms recorded contralateral and ipsilateral to the side of the synchronized visual event. The scalp topographies represent the difference between contralateral and ipsilateral activity. (b) Topographical map of the early lateralized effect based on the average ERP effect observed across the early interval identified in panel a. (c) Topographical map of the late lateralized effect based on the average ERP effect observed across the late interval identified in panel a. For display purposes, the lateral ERP scalp topography maps (N2pc and early laterality) are based on an contralateral-minus-ipsilateral difference mirrored across the vertical meridian of the electrode array with electrodes on the midline artificially set to zero (see also Hickey et al., 2009).

et al., 2000), and that has been interpreted as evidence of an enhancement of visual sensory processing (Hillyard et al., 1998; Mangun and Hillyard, 1991). The current results thus suggest that multisensory integration can enhance early visual processing. This is supported by a recent transcranial magnetic stimulation (TMS) study by Romei et al. (2009) who have shown that auditory signals enhance visual processing within early low level visual cortex (see also Romei et al., 2007).

However, this interpretation is problematic. First, as noted above, there was very little in the way of a clear visual P1 in our data. This does not preclude the possibility that multimodal integration had an impact on the residual unimodal sensory processing that occurred in the latency interval of the P1, and that this underlies the 100 ms lateralized effect evident in our data, but this suggestion would be substantially more convincing if the visual P1 could be unambiguously identified. Second, as is clear in the topographic map illustrated in Figs. 8b and 9b, the early lateralized effect has a lateral topography not generally associated with the visual P1. Thus, instead of a visual P1 we refer to this early effect as an early contralateral positivity. This effect does in any case appear to stem from strongly spatially-organized neural processing, as is evident from the fact that it is elicited contralateral to the



**Fig. 10.** Grand-averaged ERPs in the V only condition. The ERP waveforms shown here are collapsed across three posterior left and three posterior right hemisphere electrodes as in Figs. 8a and 9a. The grand averaged ERPs represent the visual ERPs collapsed over left and right visual target location and left and right hemispheres so as to yield waveforms recorded contralateral and ipsilateral to the side of the synchronized visual event. For display purposes, the lateral ERP scalp topography maps (N2pc and early laterality) are based on an contralateral-minus-ipsilateral difference mirrored across the vertical meridian of the electrode array with electrodes on the midline artificially set to zero (see also Hickey et al., 2009).

visual event. We can only speculate about the exact location of the neural generators underlying this effect. The temporo-occipital distribution across the scalp, and the contralateral organization with respect to the location of the relevant visual event, however, suggest that this effect reflects a change in the higher-order processing pathways of the visual system; presumably those processes related to the ventral pathways involved in object identification.

**Attentional selection as reflected in the N2pc (210–250 ms).** The first late lateralized interaction (210–250 ms) was further examined by separate ANOVAs for each Stimulus Type (AV<sub>target</sub>, AV<sub>distractor</sub>, vs. V<sub>only</sub>). The N2pc, as indexed by a Stimulus Location  $\times$  Hemisphere interaction, was present in the AV<sub>target</sub> condition,  $F_{1,13} = 12.19$ – $80.05$ ;  $P_s < 0.0005$ , at four consecutive 10 ms time windows (210–250 ms). In the AV<sub>distractor</sub> condition, the N2pc was evident,  $F_{1,13} = 6.97$ – $7.89$ ;  $P_s < 0.05$ , at two consecutive 10 ms time windows (210–230 ms). Thus, the early contralateral positivity was followed by an N2pc contralateral to the location of the audiovisual event (see Figs. 8c and 9c). In the visual-only condition there was no sign of an N2pc, suggesting that a target orientation change alone was not salient enough to draw attention consistently (all  $F_s < 2.47$ ;  $P_s > 0.05$ ). The ERP results thus support the conclusions we drew from behavior, namely that the sound makes the synchronized visual event pop out from its environment and draw attention.

**Lateralized VSTM effects (330–1000 ms).** The second late lateralized interaction (330–1000 ms) was further examined by a separate ANOVA with Stimulus Type (AV<sub>target</sub> vs. AV<sub>distractor</sub>), Hemisphere and Location as within subject factors. These analyses yielded a reliable interaction between 440 and 1,000 ms,  $F_{1,13} = 5.79$ – $45.29$ ;  $P_s < 0.05$ . This indicates that in the AV<sub>target</sub> condition (Fig. 8a), and to a lesser degree in the AV<sub>distractor</sub> condition (Fig. 9a), the N2pc was followed by a CNSW (Klaver et al., 1999) (also known as the CDA (Vogel and Machizawa, 2004) or SPCN (Jolicoeur et al., 2008)). The CNSW is thought to reflect neural activity related to the maintenance of information in visual short-term memory (VSTM). In our design participants were instructed to memorize the orientation of the target until the end of the entire stimulus sequence of each trial. Critically, this memorization was required only in the AV<sub>target</sub> condition. The difference in CNSW amplitude between the AV<sub>target</sub> and AV<sub>distractor</sub> conditions is thus consistent with the idea that VSTM resources were required by AV targets, but not by AV distractors.

#### Target processing as reflected in the P3 (180–660 ms)

Finally, as is clear from Fig. 6, we observed a P3 component (Sutton et al., 1965) that was larger in the AV<sub>target</sub> condition than in the A + V condition. This component was evident between 180 and 660 ms after stimulus onset,  $F_{1,13} = 5.07$ – $65.66$ ;  $P_s < 0.05$ , as indexed by a Stimulus Type (AV<sub>target</sub> vs. A + V) main effect (for correct trials). The P3 emerges in many different tasks and most likely reflects awareness of the target and/or the updating of working memory (Nieuwenhuis et al., 2005; Sutton et al., 1965).

## Discussion

This study investigated the neural mechanisms involved in shifts of attention driven by multisensory events. The main findings were that spatially uninformative auditory signals affect the competition among multiple visual objects (as indicated by behavioral measures), and that this is related to an early audiovisual interaction starting 50–60 ms after stimulus onset over parieto-occipital electrodes. Our electrophysiological results reveal a sequence of events in the information processing chain that lead to the improved detection of synchronized audiovisual events.

First, we observed an early multisensory interaction for correctly reported audiovisual targets, suggesting that rapid audiovisual

integration underlies the pip and pop phenomenon. Moreover, this early multisensory interaction was correlated with behavioral search benefits, indicating that participants with a strong multisensory interaction benefited the most from the synchronized auditory signal. The latency of this early multisensory component is consistent with earlier studies reporting low-level multisensory interactions (Giard and Peronn  t, 1999; Molholm et al., 2002; Talsma et al., 2007). For example, Lebib et al. (2003) reported a reduced P50 when auditory vowel sounds were congruent with visual mouth movements, as compared to when the vowel sound and mouth movement did not match (see also Klucharev et al., 2003; Lebib et al., 2004). More generally, there is substantial evidence that early auditory components, including the N1, are reduced when stimuli have a congruent visual component (Besle et al., 2004; Stekelenburg and Vroomen, 2007).

The multisensory interaction appears to be a sub-additive interaction, with the AV conditions showing reduced positivity relative to the A + V conditions. However, it is difficult to determine directionality of this AV interaction (see Cappe et al., 2010 for extended discussion of this issue). The polarity of ERP components does not necessarily reflect directionality of underlying neural activity and ERP activity summates at scalp surface. This means that the AV interaction identified in Fig. 6 could reflect the influence of a neural generator that becomes active in the AV conditions and expresses with negative polarity at scalp. If this co-occurred with positive polarity activity stemming from other sources, this might cause reduced overall positivity in the AV condition, and thus account for the pattern evident in Fig. 6. Importantly, this would not reflect a reduction of neural activity in the AV condition. While we cannot concretely determine if this situation underlies the current results, the data does demonstrate that the AV interaction—regardless of its direction—underlies the subsequent pip and pop effect on behavior.

The parieto-occipital distribution observed in prior studies—and in the current results—fits with claims that the inferior parietal cortex, notably the intraparietal sulcus, may be involved in early multisensory integration (Giard and Peronn  t, 1999; Molholm et al., 2002; Schroeder and Foxe, 2004). However, whereas these earlier reports point towards a somewhat right-lateralized distribution for this site, the current study found a more left-lateralized distribution. The idea that AV stimuli trigger activity in a multimodal brain area over the left temporal cortex has some basis in ERP, ERF and fMRI studies investigating bimodal speech perception (Callan et al., 2003a, 2003b; Calvert and Campbell, 2003; Miller and D'Esposito, 2005; M  tt  nen et al., 2002).

Fig. 6 suggests that the data appear to contain some residual alpha oscillations (see also Figs. 3, 8, 9 and 10). One could argue that there is a relation between the observed early multisensory effects and this pre-stimulus oscillatory activity; particularly its phase at the time of stimulus change. For instance, Mathewson et al. (2009) have shown that the detectability of a faint stimulus is strongly related to the phase of the ongoing alpha oscillation (see also Hanslmayr et al., 2007). Analogously, one might argue that our target stimuli will only be identified when presented during the most optimal phase of the alpha wave, and whether the presence of an auditory stimulus broadens this range. Although there might be a relation between alpha phase and performance, we nevertheless consider it unlikely that alpha activity alone can explain our observed early effects. (A) In the present study, the event of interest was embedded in a sequence of stimuli that each changed at irregular moments, rendering it unlikely that our visual sequence evoked a strong train of alpha waves. (B) We report a reliable correlation across participants between the benefits in the AV<sub>target</sub> condition (relative to the V condition) and magnitude of the early multisensory interaction, indicating that those participants showing a strong multisensory interaction effect benefited the most from the audiovisual targets (see Fig. 7). Therefore, we consider it as unlikely that one participant was differently affected by alpha-phase than another. Thus, the presence of the correlation bolsters our claim that the multisensory interaction is important for the pip and pop effect.

Nevertheless, the possible relation between alpha phase, stimulus detectability and a possible influence of auditory stimulation on this process is interesting and deserves further study in future research.

The early integration of audiovisual events appears to have an impact on visual processing, as is evident in modulation of posterior activity over cortex contralateral to the visual aspect of the AV event (see Figs. 8 and 9). The latency (80–100 ms) of this effect is characteristic of a modulation of the visual P1 component, and the amplitude of the contralateral P1 is typically larger when sensory processing is facilitated (Busse et al., 2005; Heinze et al., 1994; Hillyard et al., 1998; Luck et al., 2000; Talsma and Woldorff, 2005b). P1 effects such as these are thought to reflect an increase in the gain of sensory activity (Hillyard et al., 1998; Mangun and Hillyard, 1991). One interpretation of this effect is thus that it reflects an increase in sensory or perceptual activity in visual cortex, reflecting an increase in visual salience (see e.g. Romei et al., 2009, 2007). This would be consistent with the behavioral and N2pc data demonstrating that the AV target pops out from its cluttered environment.

However, it is important to note that the interpretation of the early lateralized effect as a modulation of a *typical* visual P1 is problematic, since the topographical maps show a more lateral, temporal-occipital distribution, and there is little in the way of a standard P1 in the ERP waveforms. Therefore, we prefer to refer to this effect as an early contralateral positivity instead. It should be noted that there was not a very strong exogenous P1 component present in our data and we believe that this absence is largely due to our experimental design: target events (to which the ERPs were time-locked) were embedded in a continuous stream of very similar distractor events, and these types of dynamic displays are known to attenuate early exogenous aspects of the visual ERP (e.g. Leblanc et al., 2008; Martens et al., 2007). Furthermore, in our experimental design the visual P1 elicited by an AV event would co-occur and be superimposed upon a volume conducted exogenous auditory N1. However, neither of these possibilities explains the atypical topography of the P1-latency effect (see Figs. 8a and 9a). The results suggest nevertheless that this effect stems from a spatially-organized neural mechanism, because the effect is observed contralateral to the visual aspect of the AV events. Although further research is clearly needed to determine the degree to which this effect reflects a change in visual processing, we hypothesize that this effect is compatible with a multisensory driven change in sensitivity of the higher-order (i.e. ventral pathway) visual processing stream. The latency of the early contralateral positivity is consistent with other studies showing activity in the ventral pathway during this time-window (Kirchner and Thorpe, 2006; Lamme and Roelfsema, 2000).

The pattern of N2pc results is in line with the idea that the cross-modal boost of visual processing results in selection of the audiovisual event. The visual object is subsequently processed as a behaviorally relevant stimulus, resulting in an enhanced P3. If it is a target, the response-related property is maintained in visual short term memory, resulting in an increase in CNSW amplitude. The overall pattern of ERP results is thus consistent with our understanding of the cognitive operations at work in multimodal perception and cognition.

Whereas other studies report that attention can affect multisensory processing (Alsius et al., 2005, 2007; Busse et al., 2005; Fairhall and Macaluso, 2009; Senkowski et al., 2008; Talsma et al., 2007; Talsma and Woldorff, 2005b; Van Ee et al., 2009), the present study demonstrates for the first time that multisensory processing can affect attention. Our data suggest that multisensory integration can occur pre-attentively and that some of these integration mechanisms are not influenced by top-down control processes based on the relevance of the audiovisual event (Bertelson et al., 2000; Driver, 1996). We note, however, that we do not believe that the entire process is immune to top-down effects. For example, Vroomen and de Gelder (2000) have demonstrated that auditory signals affect the processing of a synchronized event when they used a salient auditory event, but not when the auditory event was part

of a melody, suggesting that temporal synchrony alone does not necessarily result in multisensory integration (see also Van der Burg et al., 2010b). It appears that the auditory stimulus needs to be a marked event in order for it to be uniquely bound to a visual event. This may also explain why attention to at least one modality (Van der Burg et al., 2010a), or both modalities (Talsma et al., 2007) was a prerequisite for early multisensory integration. In the Talsma et al. study the auditory and visual events were part of a sequence, and were not salient events in themselves. We propose here that when participants are in a distributed state of attention and auditory events are relatively rare, audiovisual integration occurs not only easily and automatically, but can even serve the purpose of guiding attention to the location of the concurrently presented visual stimulus. This notion is consistent with a recent framework proposed by Talsma et al. (2010), which states that top-down controlled voluntary attention is predominantly necessary in situations where none of the individual stimuli are salient enough by themselves.

## Summary and conclusion

The present study employed a visual search paradigm, using complex and dynamically changing visual displays and brief auditory events to show how sounds can affect the competition among visual stimuli. We found a systems-level cascading sequence of event that started with an early latency multisensory integration effect. This integration effect ultimately led to a rapid shift of attention to a visual stimulus that was presented in temporal synchrony with the occurrence of a transient auditory event. While other studies have reported early effects of audition on vision when using single audiovisual events (Giard and Peronn  t, 1999; Molholm et al., 2002; Talsma et al., 2007), the present study is the first to reveal how such early effects contribute to biasing or resolving the competition between multiple visual events. In this respect it complements earlier studies showing how vision can aid audition in noisy environments (Senkowski et al., 2008). This speaks to the importance of multisensory processing in real life, where we are exposed to an overwhelming amount of information from different senses. In this highly dynamic environment, synchronized information from two different senses appears to be more than just the sum of what is provided by either sense individually.

## Acknowledgments

We wish to thank Micah Murray and an anonymous reviewer for their suggestion to correlate the early AV effect with behavior in the pip and pop task.

DT and CNLO contributed equally to this project and should both be considered as second author. This research was supported by a Dutch Technology Foundation STW grant (07079), a division of NWO and the Technology Program of the Ministry of Economic Affairs (to JT), and a NWO-VIDI grant 452-06-007 (to CNLO).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.neuroimage.2010.12.068.

## References

- Alais, D., Burr, D., 2004. The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262.
- Alsius, A., Navarra, J., Campbell, R., Soto-Faraco, S., 2005. Audiovisual integration of speech falters under attention demands. *Curr. Biol.* 15, 839–843.
- Alsius, A., Navarra, J., Soto-Faraco, S., 2007. Attention to touch weakens audiovisual speech integration. *Exp. Brain Res.* 183, 399–404.
- Bertelson, P., Vroomen, J., De Gelder, B., Driver, J., 2000. The ventriloquist effect does not depend on the direction of deliberate visual attention. *Percept. Psychophys.* 62, 321–332.



- Besle, J., Fort, A., Delpuech, C., Giard, M.H., 2004. Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur. J. Neurosci.* 20, 2225–2234.
- 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.* 102, 18751–18756.
- Callan, D.E., Callan, A.M., Kroos, C., Vatikiotis-Bateson, E., 2003a. Neural processes underlying perceptual enhancement by visual speech gestures. *Cogn. Brain Res.* 10, 349–353.
- Callan, D.E., Jones, J.A., Muhall, K., Callan, A.M., Kroos, C., Vatikiotis-Bateson, E., 2003b. Neural processes underlying perceptual enhancement by visual speech gestures. *NeuroReport* 2213–2218.
- Calvert, G., Campbell, A.J., 2003. Reading Speech from Still and Moving Faces: The Neural Substrates of Visible Speech. *J. Cogn. Neurosci.* 15, 57–70.
- Calvert, G.A., Campbell, R., Brammer, M.J., 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10, 649–657.
- Cappe, C., Thut, G., Romei, V., Murray, M.M., 2010. Auditory-visual multisensory interactions in humans: Timing, topography, Directionality, and sources. *J. Neurosci.* 30, 12572–12580.
- Chen, Y.C., Yeh, S.L., 2009. Catch the moment: multisensory enhancement of rapid visual events by sound. *Exp. Brain Res.* 198, 209–219.
- Driver, J., 1996. Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature* 381, 66–68.
- Eckert, M.A., Kamdar, N.V., Chang, C.E., Beckmann, C.F., Greicius, M.D., Menon, V., 2008. A Cross-Modal System Linking Primary Auditory and Visual Cortices: Evidence From Intrinsic fMRI Connectivity Analysis. *Hum. Brain Mapp.* 29, 848–857.
- Fairhall, S.L., Macaluso, E., 2009. Spatial attention can modulate audiovisual integration at multiple cortical and subcortical areas. *Eur. J. Neurosci.* 29, 1247–1257.
- Giard, M.H., Peronn  t, F., 1999. Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysical study. *J. Cogn. Neurosci.* 11, 473–490.
- Hanslmayr, S., Aslan, A., Staudigi, T., LKlimesch, W., Herrmann, C.S., Baumi, K.-H., 2007. Prestimulus oscillations predict visual perception performance between and within subjects. *NeuroImage* 37, 1465–1473.
- Heinze, H.J., Mangun, G.R., Burchert, W., Hinrichs, H., Scholz, M., M  nte, T.F., G  s, A., Scherg, M., Johannes, S., Hundeshagen, H., Gazzaniga, M.S., Hillyard, S.A., 1994. Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature* 372, 543–546.
- Hershenson, M., 1962. Reaction time as a measure of intersensory facilitation. *J. Exp. Psychol.* 63, 289–293.
- Hickey, C., Di Lollo, V., McDonald, J.J., 2009. Electrophysiological indices of target and distractor processing in visual search. *J. Cogn. Neurosci.* 18, 604–613.
- Hillyard, S.A., Vogel, E.K., Luck, S.J., 1998. Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philos. Trans. R. Soc. B Biol. Sci.* 353, 1257–1270.
- Jolicoeur, P., Brisson, B., Robitaille, N., 2008. Dissociation of the N2pc and sustained posterior contralateral negativity in a choice response task. *Brain Res.* 1215, 160–172.
- Kirchner, H., Thorpe, S.J., 2006. Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vis. Res.* 46, 1762–1776.
- Klaver, P., Talsma, D., Wijers, A.A., Heinze, H.J., Mulder, G., 1999. An event-related brain potential correlate of visual short-term memory. *NeuroReport* 10, 2001–2005.
- Klucharev, V., M  tt  nen, R., Sams, M., 2003. Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cogn. Brain Res.* 18, 65–75.
- Lamme, V.A.F., Roelfsema, P.R., 2000. the distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579.
- Lebib, R., Papo, D., de Bode, S., M.B.P., 2003. Evidence of a visual-to-auditory cross-modal sensory gating phenomenon as reflected by the human P50 event-related brain potential modulation. *Neurosci. Lett.* 8, 185–188.
- Lebib, R., Papo, D., Douiri, A., De Bode, S., Gillon Dowens, M., Baudonni  re, P.M., 2004. Modulations of 'late' event-related brain potentials in humans by dynamic audiovisual speech stimuli. *Neurosci. Lett.* 372, 74–79.
- Leblanc, E., Prime, D.J., Jolicoeur, P., 2008. Tracking the location of visuospatial attention in a contingent capture paradigm. *J. Cogn. Neurosci.* 20, 657–671.
- Luck, S.J., Hillyard, S.A., 1994. Spatial filtering during visual search: evidence from human electrophysiology. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 1000–1014.
- Luck, S.J., Woodman, G.F., Vogel, E.K., 2000. Event-related potential studies of attention. *Trends Cogn. Sci.* 4, 432–440.
- Macaluso, E., Frith, C.D., Driver, J., 2000. Modulation of human visual cortex by cross-modal spatial attention. *Science* 289, 1206–1208.
- Mangun, G.R., Hillyard, S.A., 1991. Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual spatial priming. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 1057–1074.
- Martens, S., Munneke, J., Smid, H., Johnson, A., 2007. Quick minds don't blink: Electrophysiological correlates of individual differences in attentional selection. *J. Cogn. Neurosci.* 18, 1423–1438.
- Martuzzi, R., Murray, M.M., Michel, C.M., Thiran, J.P., Maeder, P.P., Clarke, S., Meuli, R.A., 2007. Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cereb. Cortex* 17, 1672–1679.
- Mathewson, K.E., Gratton, G., Fabiani, M., Beck, D.M., Ro, T., 2009. To see or not to see: Pre-stimulus alpha phase predicts visual awareness. *J. Neurosci.* 29, 2725–2732.
- McDonald, J.J., Teder-S  lejarvi, W.A., Di Russo, F., Hillyard, S.A., 2005. Neural basis of auditory-induced shifts in visual time-order perception. *Nat. Neurosci.* 8, 1197–1202.
- McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. *Nature* 264, 746–748.
- Miller, B.T., D'Esposito, M., 2005. Searching for "the top" in top-down control. *Neuron* 48, 535–538.
- Mishra, J., Martinez, A., Sejnowski, T.J., Hillyard, S.A., 2007. Early Cross-Modal Interactions in Auditory and Visual Cortex Underlie a Sound-Induced Visual Illusion. *J. Neurosci.* 27, 4120–4131.
- Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., Foxe, J.J., 2002. Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Brain Res. Cogn. Brain Res.* 14, 115–128.
- M  tt  nen, R., Krause, C., Tipples, K., Sams, M., 2002. Processing of changes in visual speech in the human auditory cortex. *Cogn. Brain Res.* 13.
- Ngo, M.K., Spence, C., 2010. Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Atten. Percept. Psychophys.* 72, 1654–1665.
- Nieuwenhuis, S., Aston-Jones, G., Cohen, J.D., 2005. Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychol. Bull.* 131, 510–532.
- Olivers, C.N.L., Van der Burg, E., 2008. Bleeping you out of the blink: Sound saves vision from oblivion. *Brain Res.* 1242, 191–199.
- Romei, V., Murray, M.M., Merabet, L.B., Thut, G., 2007. Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: Implications for multisensory interactions. *J. Neurosci.* 27, 11465–11472.
- Romei, V., Murray, M.M., Cappe, C., Thut, G., 2009. Preperceptual and Stimulus-Selective Enhancement of Low-Level Human Visual Cortex Excitability by Sounds Current Biology 19, 1799–1805.
- Schroeder, C.E., Foxe, J.J., 2004. Multisensory convergence in early cortical processing. In: Calvert, G.A., Spence, C., Stein, B.E. (Eds.), *The handbook of multisensory processes*. MIT press, New York, pp. 295–309.
- Schroeder, C.E., Foxe, J.J., 2005. Multisensory contributions to low-level, 'unisensory' processing. *Curr. Opin. Neurobiol.* 15, 454–458.
- Senkowski, D., Saint-Amour, D., Gruber, T., Foxe, J.J., 2008. Look who's talking: The deployment of visuo-spatial attention during multisensory speech processing under noisy environmental conditions. *NeuroImage* 43, 379–387.
- Shipley, T., 1964. Auditory flutter-driving of visual flicker. *Science* 145, 1328–1330.
- Spence, C., 2007. Audiovisual multisensory integration. *Acoust. Sci. Technol.* 28, 61–70.
- Stein, B.E., Meredith, M.A., 1993. *The merging of the senses*. MIT press, Cambridge, MA.
- Stein, B.E., London, N., Wilkinson, L.K., Price, D.D., 1996. Enhancement of perceived visual intensity by auditory stimuli: A psychological analysis. *J. Cogn. Neurosci.* 8, 497–506.
- Stekelenburg, J.J., Vroomen, J., 2007. correlates of multisensory integration of ecologically valid audiovisual events. *J. Cogn. Neurosci.* 19, 1964–1973.
- St  rmer, V.S., McDonald, J.J., Hillyard, S.A., 2009. Cross-modal cueing of attention alters appearance and early cortical processing of visual stimuli. *Proc. Natl Acad. Sci.* 106, 22456–22461.
- Sutton, S., Braren, M., Zubin, J., John, E.R., 1965. Evoked-potential correlates of stimulus uncertainty. *Science* 150, 1187–1188.
- Talsma, D., 2008. Auto-adaptive averaging: Detecting artifacts in event-related potential data using a fully automated procedure. *Psychophysiology* 45, 216–228.
- Talsma, D., Woldorff, M.G., 2005a. Methods for estimation and removal of artifacts and overlap in ERP data. In: Handy, T. (Ed.), *Event-Related Potentials: A Methods Handbook*. MIT Press, Cambridge, MA, pp. 115–148.
- Talsma, D., Woldorff, M.G., 2005b. Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *J. Cogn. Neurosci.* 12, 1098–1114.
- Talsma, D., Doty, T.J., Woldorff, M.G., 2007. Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cereb. Cortex* 17, 691–701.
- Talsma, D., Senkowski, D., Soto-Faraco, S., Woldorff, M.G., 2010. The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci.* 14, 400–410.
- Teder-S  lejarvi, W.A., McDonald, J.J., Di Russo, F., Hillyard, S.A., 2002. An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cogn. Brain Res.* 14, 106–114.
- Theeuwes, J., 2000. Attentional capture and oculomotor control. *Perception* 29 (Supplement, 2a).
- Van der Burg, E., Olivers, C.N.L., Bronkhorst, A.W., Theeuwes, J., 2008a. Audiovisual events capture attention: Evidence from temporal order judgments. *J. Vis.* 8 (art: 2).
- Van der Burg, E., Olivers, C.N.L., Bronkhorst, A.W., Theeuwes, J., 2008b. Pip and pop: Non-spatial auditory signals improve spatial visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1053–1065.
- Van der Burg, E., Olivers, C.N.L., Bronkhorst, A.W., Theeuwes, J., 2009. Poke and pop: Tactile-visual synchrony increases visual saliency. *Neurosci. Lett.* 450, 60–64.
- Van der Burg, E., Brederoo, S.G., Nieuwenstein, M.R., Theeuwes, J., Olivers, C.N.L., 2010a. Audiovisual semantic interference and attention: Evidence from the attentional blink paradigm. *Acta Psychol.* 134, 198–205.
- Van der Burg, E., Cass, J., Olivers, C.N.L., Theeuwes, J., Alais, D., 2010b. Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS ONE* 5, e10664.
- Van Ee, R., Van Boxtel, J.J.A., Parker, A.M., Alais, D., 2009. Multisensory Congruency as a Mechanism for Attentional Control over Perceptual Selection. *J. Neurosci.* 29, 11641–11649.
- Vogel, E.K., Machizawa, M.G., 2004. Neural activity predicts individual differences in visual working memory capacity. *Nature* 428, 748–751.
- Vroomen, J., De Gelder, B., 2000. Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 1583–1590.
- Zangenehpour, S., Zatorre, R.J., 2010. Crossmodal recruitment of primary visual cortex following brief exposure to bimodal audiovisual stimuli. *Neuropsychologia* 48, 591–600.