

# Exogenous spatial attention decreases audiovisual integration

N. Van der Stoep · S. Van der Stigchel · T. C. W. Nijboer

Published online: 24 October 2014  
© The Psychonomic Society, Inc. 2014

**Abstract** Multisensory integration (MSI) and spatial attention are both mechanisms through which the processing of sensory information can be facilitated. Studies on the interaction between spatial attention and MSI have mainly focused on the interaction between endogenous spatial attention and MSI. Most of these studies have shown that *endogenously* attending a multisensory target enhances MSI. It is currently unclear, however, whether and how *exogenous* spatial attention and MSI interact. In the current study, we investigated the interaction between these two important bottom-up processes in two experiments. In Experiment 1 the target location was task-relevant, and in Experiment 2 the target location was task-irrelevant. Valid or invalid exogenous auditory cues were presented before the onset of unimodal auditory, unimodal visual, and audiovisual targets. We observed reliable cueing effects and multisensory response enhancement in both experiments. To examine whether audiovisual integration was influenced by exogenous spatial attention, the amount of race model violation was compared between exogenously attended and unattended targets. In both Experiment 1 and Experiment 2, a *decrease* in MSI was observed when audiovisual targets were exogenously attended, compared to when they were not. The interaction between exogenous attention and MSI was less pronounced in Experiment 2. Therefore, our results indicate that exogenous attention diminishes MSI when spatial

orienting is relevant. The results are discussed in terms of models of multisensory integration and attention.

**Keywords** Cueing · Exogenous · Bottom-up · Attention · Multisensory integration

## Introduction

Two processes that are involved in the interaction between information from different senses are multisensory integration (MSI) and crossmodal attention. Both MSI and (crossmodal) attention are able to facilitate the speed of detection, and the accuracy of localization and identification of targets (e.g., Leo, Bolognini, Passamonti, Stein, & Ladavas, 2008; Montagna, Pestilli, & Carrasco, 2009; Spence & Driver, 1997; Stevenson, Krueger Fister, Barnett, Nidiffer, & Wallace, 2012). To date, however, it is unclear under what circumstances and how these two processes interact. Although some studies have found that MSI occurs independent of whether attention has been allocated to the multisensory stimulus (e.g., Bertelson, Vroomen, de Gelder, & Driver, 2000; Bertelson, Pavani, Ladavas, Vroomen, & de Gelder, 2000; Soto-Faraco, Navarra, & Alsius, 2004; Vroomen, Bertelson, & de Gelder, 2001), other studies have shown that attention is able to modulate MSI (e.g., Alsius, Navarra, & Soto-Faraco, 2007; Fairhall & Macaluso, 2009; Talsma & Woldorff, 2005; Talsma, Doty, & Woldorff, 2007). To explain these different findings, it has been suggested that the influence of attention on MSI depends on several factors such as the type of task (e.g., detection vs. identification), the stimulus properties (e.g., salient vs. near threshold, simple vs. complex), and the attentional resources that are available (e.g., low attentional load vs. high attention load; exogenous vs. endogenous attention manipulation; for reviews see Koelewijn, Bronkhorst, &

---

N. Van der Stoep (✉) · S. Van der Stigchel · T. C. W. Nijboer  
Department of Experimental Psychology, Helmholtz Institute,  
Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The  
Netherlands  
e-mail: n.vanderstoep@uu.nl

T. C. W. Nijboer  
Brain Center Rudolf Magnus, and Center of Excellence for  
Rehabilitation Medicine, University Medical Center Utrecht and De  
Hoogstraat Rehabilitation, Utrecht, The Netherlands

Theeuwes, 2010; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010).

Interestingly, studies of the interaction between attention and MSI have mainly looked into the influence of endogenous attention on MSI. The results from these studies indicate that endogenous attention influences MSI, either by showing that it enhances MSI (e.g., Talsma & Woldorff, 2005; Fairhall & Macaluso, 2009; but see Bertelson, Vroomen, de Gelder, & Driver, 2000), that MSI is decreased when attentional resources are limited or depleted (Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Alsius, Navarra, & Soto-Faraco, 2007), and that endogenous spatial attention spreads across the unimodal components of a multisensory stimulus (Busse et al., 2005). In contrast with these findings, Zou, Müller, and Shi (2012) observed larger benefits of multisensory stimulation on visual target detection in the endogenously *unattended* half of the stimulus display as compared to the endogenously *attended* half of the display (as evident from a larger pip-and-pop effect at the unattended side). A factor that might have contributed to these different findings is the degree of spatial uncertainty of the target in the task. For example, when the target location was not varied on a trial-by-trial basis, attention was shown to enhance multisensory integration for simple (e.g., Talsma & Woldorff, 2005) and complex stimuli (e.g., Alsius et al., 2005, 2007; Fairhall & Macaluso, 2009).

Differences in the specific task requirements and stimulus properties aside (e.g., see Navarra, Alsius, Soto-Faraco, & Spence, 2010, for a discussion), the majority of these studies have at least shown that MSI *can* be influenced by endogenous forms of attention. Much less is known about the effects of exogenous spatial attention on MSI. In fact, there is some debate about whether crossmodal exogenous spatial attention and MSI are actually different processes or the same process (e.g., Macaluso, Frith, & Driver, 2000; McDonald, Teder-Sälejärvi, & Ward, 2001; Spence, 2010, pp. 183–184). Some researchers have suggested that exogenous spatial attention and multisensory integration can be discriminated, for example, based on the time-course of their effects. The effects of exogenous spatial attention are often most pronounced with cue target onset asynchronies (CTOA) of 50–200 ms (e.g., Berger, Henik, & Rafal, 2005; Spence & Driver, 1994), whereas multisensory integration is typically most pronounced for stimuli that are presented in close temporal proximity (SOAs < ~100 ms, e.g., Leone & McCourt, 2013; Meredith, Nemitz, & Stein, 1987; Stevenson, Krueger Fister, Barnett, Nidiffer, & Wallace, 2012). Observations of a temporal binding window (TBW) in behavioral studies of temporal order judgment and simultaneity judgment are in line with such a window of integration (~100 visual lead and ~60 auditory lead, but note that the width of this window is task and stimulus dependent, see for example Hirsh & Sherrick, 1961; Keetels & Vroomen, 2005; Stevenson & Wallace, 2013;

Vroomen & Keetels, 2010; Zampini, Guest, Shore, & Spence, 2005). Whereas a distinction between crossmodal exogenous attention and multisensory integration based on differences in temporal properties seems to hold well based on the behavioral findings in the literature, others have argued that this distinction is somewhat problematic in terms of the underlying neural interactions given that some researchers have reported observations of multisensory integration in multisensory neurons with SOAs larger than 100–200 ms (see McDonald, Teder-Sälejärvi, & Ward, 2001, for a discussion).

Whereas there is little research on whether exogenous spatial attention influences multisensory integration, there are several studies of whether the benefits of spatial attention shifts that are evoked by multisensory exogenous spatial cues are any different from unimodal exogenous cues (i.e., the effect of multisensory integration on exogenous spatial attention; e.g., Santangelo, Van der Lubbe, Belardinelli, & Postma, 2006; Santangelo, Ho, & Spence, 2008; Santangelo, Van der Lubbe, Belardinelli, and Postma 2008). Under low cognitive load, the size of the cueing effect (i.e., response times (RTs) validly cued targets < RTs invalidly cued targets) does not differ between multisensory and unimodal exogenous spatial cues. Under high cognitive load (i.e., while performing a secondary task), however, multisensory exogenous cues are the only cues that are able to evoke a cueing effect whereas unimodal cues do not evoke a significant cueing effect anymore (Santangelo, Ho, & Spence, 2008; see Spence & Santangelo, 2009, for a review). Other studies on multisensory processing and attention, which have mainly focused on temporal stimulus properties, have also observed benefits of multisensory stimulation, but in the detection of visual targets embedded in a complex visual environment (e.g., the freezing phenomenon, Vroomen & De Gelder, 2000; the pip and pop effect, Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008; Van der Burg, Talsma, Olivers, Hickey, & Theeuwes, 2011; see Ngo & Spence, 2010, for influences of both spatial and temporal alignment on multisensory enhancement of visual search). In the study by Van der Burg et al. (2011) the benefits of multisensory stimulation were most pronounced when the auditory cue and the visual target were presented simultaneously. These benefits could not be explained by a general alerting effect or by shifts of spatial attention (the sound was not lateralized, but presented to the left and right ear through headphones). These findings indicate that the observed benefits were probably due to the attention attracting effects of the integration of auditory and visual stimuli and are in line with a distinction between crossmodal attention and multisensory integration based on time differences. Although these studies provide insight into the influence of multisensory integration on possible shifts of exogenous spatial attention (and some also on whether MSI can occur pre-attentively, e.g., Soto-Faraco, Navarra, & Alsius, 2004; Spence & Driver, 2000; Vroomen, Bertelson, and De Gelder, 2001), they do not

provide information on the influence of exogenous spatial attention on the integration of information from multiple senses at exogenously attended compared to exogenously unattended locations.

One study that did look into the influence of exogenous spatial attention on MSI was performed by Vroomen, Bertelson, and De Gelder (2001). In their study they investigated whether the ventriloquist effect (which is thought to be the result of MSI) was affected by the direction of exogenous spatial attention. They did not find such an influence of exogenous spatial attention, suggesting that exogenous spatial attention is not able to influence MSI. Yet, the conditions for observing an influence of exogenous spatial attention on MSI may not have been optimal in their study. Their main experiment consisted of the simultaneous presentation of four squares and a sound that had to be localized by participants (“was the sound coming from the left or the right?”). One of the squares was smaller than the other three, which caused it to act as a singleton and therefore automatically attracted attention. The exogenous cue (i.e., a singleton) was presented simultaneously with the auditory stimulus and was part of the bimodal stimulus. Their results showed that the ventriloquist effect did not depend on the direction of exogenous spatial attention. Several studies, however, have shown that it takes some time for exogenous spatial attention to develop its strongest effect (e.g., 100–300 ms, see Driver & Spence, 1998 for a review on exogenous and endogenous crossmodal spatial attention). Therefore, the onset of the integration process and the shift of exogenous spatial attention may have been temporally misaligned. This may have resulted in pre-attentive MSI (i.e., no influence of exogenous spatial attention on MSI), just because there was not enough time for exogenous attention to be shifted to the location of the cue.

In order to determine whether exogenous spatial attention is able to influence MSI, we investigated this interaction in a situation in which the exogenous cue was not only presented prior to the multisensory stimulus but was also not part of the multisensory target, using simple stimuli (sounds and light disks). This ensured that there was enough time for exogenous spatial attention to be allocated to the location of the multisensory stimulus. Additionally, as the stimulus that is causing an exogenous shift of spatial attention is different from the stimuli that need to be integrated, exogenous orienting of attention and MSI do not depend on the same stimulus, providing an opportunity for both processes to emerge and for exogenous spatial attention to influence MSI. Although in several studies it was shown that endogenous attention increases multisensory integration, there are also some studies that seem to suggest that multisensory integration is enhanced when the location of the multisensory stimulus is unattended. For example, multisensory integration is less affected by the depletion of attentional resources (e.g., Santangelo, Ho, & Spence, 2008) and multisensory benefits during visual search

are larger at endogenously unattended regions of space (Zou, Müller, & Shi, 2012).

We hypothesized that if the interaction between exogenous attention and MSI depends on whether there was enough time for exogenous spatial attention to be allocated, then an interaction between exogenous spatial attention and MSI should be observed if the exogenous cue is presented slightly before the audiovisual stimulus. If this is indeed the case, two different results may be expected based on previous studies. On the one hand, exogenous attention may show an enhancement of MSI just as endogenous attention has been shown to enhance MSI (e.g., Ngo & Spence, 2010; Talsma & Woldorff, 2005). That is, multisensory integration is enhanced at exogenously attended as compared to exogenously unattended target locations. On the other hand, one might expect multisensory integration to be more pronounced at exogenously unattended locations, as several studies have shown that multisensory integration is more pronounced at endogenously unattended locations (e.g., Zou, Müller, & Shi, 2012) and that multisensory cues are less affected by the depletion of attentional resources as compared to unimodal cues (e.g., Santangelo, Ho, & Spence, 2008, Spence & Santangelo, 2009).

## Experiment 1

### Materials and methods

#### Participants

Sixteen healthy participants were tested in the experiment (seven male, nine female, mean age = 25.70 years,  $SD = 3.08$ ). This sample size was based on previous studies on exogenous spatial attention and on studies on the interaction between exogenous and endogenous attention and MSI in which sample sizes varied between 12 and 20 participants (e.g., Spence & Driver, 1997; Talsma & Woldorff, 2005; Vroomen, Bertelson, & De Gelder, 2001). Participants had normal or corrected-to-normal visual acuity, did not report any hearing problems, and received course credits for their participation. The experiment was conducted in accordance with the Declaration of Helsinki, and participants signed informed consent before the start of the experiment.

#### Apparatus

An Acer X1261P projector (60 Hz) was used to project visual stimuli on a black projection screen ( $50 \times 75$  cm). The screen was placed at 87 cm in front of the participant, whose head was placed in a chin-rest. Three speakers (Harman/Kardon HK206, Frequency response: 90–20,000 Hz) were used to present the auditory stimuli.

### Stimuli, task, and procedure

The participants had to detect visual, auditory, and audiovisual targets to the left or the right of a fixation cross (black plus sign,  $0.8^\circ \times 0.8^\circ$ ,  $0.9 \text{ cd/m}^2$  as measured with a PhotoResearch SpectraScan PR 650 spectrometer, Weber contrast  $-1.0$ ) with and without the prior presentation of an exogenous auditory cue. Target modality (visual, auditory, audiovisual) was randomized and we used the implicit cueing paradigm to make space relevant for the task to increase the possibility of finding a validity effect of the exogenous spatial cue (Ward, McDonald, & Lin, 2000). There were three possible cue and target locations: left peripheral, center, and right peripheral. Cues and targets were presented at the three locations each with equal probability: 33 % left, 33 % center, and 33 % right. Participants were instructed to keep fixating on the fixation cross throughout the experiment and to press a button as quickly as possible when a target was presented to the left or the right of the central fixation cross (i.e., a Go trial), but not when a target was presented at the center location (i.e., a No-go trial). There was only one button to respond to the presence of a target, avoiding response-priming effects while the spatial location of the target still remained relevant for the task. Response-priming effects typically occur when participants have to indicate a side where a target is being presented using two buttons: one button for targets presented on the left, and another button for targets presented on the right. If cues and targets are presented from the left and the right side, then faster responses to validly cued targets as compared to invalidly cued targets can also be explained by the priming of the side that should be responded to and/or the hand that should be responded with. Here, we avoided response priming by using only one button to respond (e.g., Ward, McDonald, & Lin, 2000).

The experiment consisted of two blocks: one block without auditory exogenous cues and one block with auditory exogenous cues. Response times in the block without cues acted as a ‘baseline’ measure of audiovisual integration, and were always presented first. The second block was always the block in which cues were present. A direct comparison of MSI between the No Cue and Cue conditions would be problematic because of possible differences in training and/or fatigue, and the lack of an alerting signal and a temporal warning signal in the No Cue condition. The comparison of MSI in the No Cue block and the Cue block was not the focus of the current study, but rather a comparison of MSI between the Valid Cue and the Invalid Cue condition. We did incorporate the No Cue block in order to know whether any race model violations could be observed using the current stimulus parameters and paradigm. In the second block, cues were present during each trial and were presented with equal probability at the three locations (left, center, right).

Each block’s visual, auditory, and audiovisual targets were randomly presented at one of three locations with equal

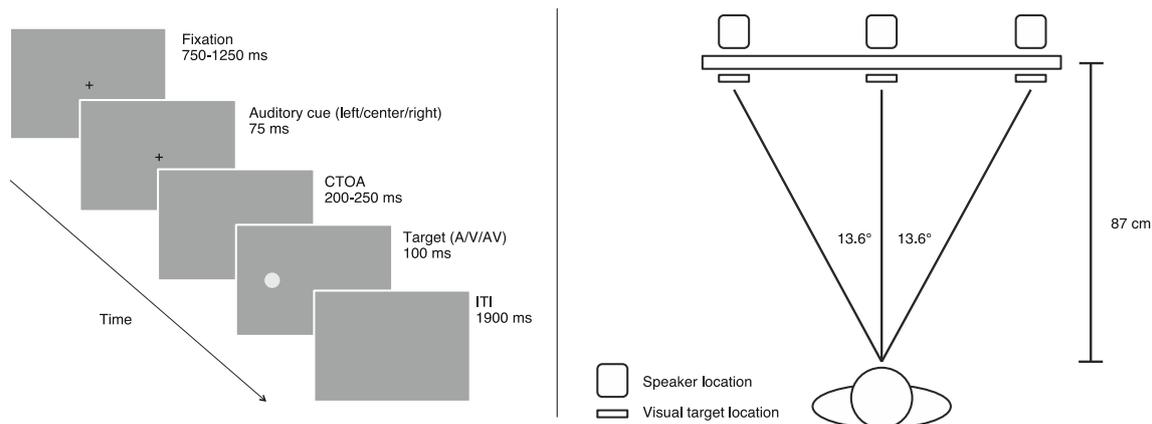
probability either without (during the first block) or with the prior presentation of an auditory exogenous cue (during the second block). The first block (No Cue) consisted of 120 go trials (40 auditory, 40 visual, and 40 audiovisual go trials) and 60 No-Go trials (20 auditory, 20 visual, and 20 audiovisual No-Go trials). The second block (cue present) consisted of 360 go trials: 120 Valid Cue trials (cue presented at the same location as the target; 40 auditory, 40 visual, and 40 audiovisual targets), 120 Invalid Cue trials (cue presented at the opposite location as the target; 40 auditory, 40 visual, and 40 audiovisual targets), and 120 Center Cue trials (cue presented at the central location; 40 auditory, 40 visual, and 40 audiovisual targets). There were 180 No-Go trials in the second block, containing 60 left cue trials, 60 center cue trials, and 60 right cue trials, with each Cue Type containing 20 auditory, 20 visual, and 20 audiovisual target stimuli that were presented at the center location.

Targets and cues were presented at eye-height at  $13.6^\circ$  degrees of visual angle to the left and the right of the fixation cross, and at the location of the fixation cross (directly in front of the participant). Visual targets consisted of a white filled circle ( $2.8^\circ \times 2.8^\circ$ ,  $245 \text{ cd/m}^2$ , Weber contrast 1.47) and were presented on a grey background ( $99.2 \text{ cd/m}^2$ ). Auditory targets consisted of a 100-ms white noise burst of  $\pm 70 \text{ dB(A)}$  SPL (as measured with an audiometer from the location of the participant; 15-ms rise and fall of the signal). Audiovisual targets consisted of a combination of the auditory and visual stimulus and were always spatially and temporally aligned (timing was confirmed with an oscilloscope). Auditory cues consisted of a 75-ms, 2000-Hz sine wave of  $\pm 78 \text{ dB}$  (15-ms rise and fall of the signal) that were presented with random cue target onset asynchronies between 200 and 250 ms (with steps of 16.7 ms). Each speaker was placed directly behind the projection screen and the center of each speaker was horizontally and vertically aligned with the location of projected visual targets. A schematic top view of the experimental setup is shown in the right panel of Fig. 1. The setup was placed in the center of the room in order to keep the auditory reflections as similar as possible for sounds presented to the left and to the right of the fixation cross.

Each trial started with the presentation of a fixation cross with a random duration between 750 and 1250 ms. After this, the fixation cross disappeared and the auditory cue was presented for 75 ms, followed by the presentation of a 100 ms auditory, visual, or audiovisual target after a random cue target onset asynchrony between 200 and 250 ms. Participants were able to respond until the end of the inter-trial interval (ITI) of 1900 ms starting at target offset. The ITI consisted of the background only. A schematic overview of the procedure is shown in the left panel of Fig. 1.

### Data analysis

*Pre-processing* Incorrect trials and trials with response times shorter than 100 ms or longer than 1000 ms were removed



**Fig. 1** A schematic overview of the procedure (left panel) and a schematic top view of the setup (right panel)

from further analysis as they were assumed to be the result of either anticipation or not paying attention to the task, respectively. This led to the removal of 2.9 % of the data. In the No Cue condition 0.2 % of the Go trials (anticipations and slow responses) and 7.8 % of the No-go trials (False Alarms) were removed, and 0.5 % of the Go trials (anticipations and slow responses) and 7.9 % of the No-go trials (False Alarms) in the cued conditions. One participant was removed from further analyses and replaced with a new participant because of low accuracy on catch trials in the No-Go No Cue condition (60 % accuracy). Median response times of each participant in each condition were used for further analyses.

For each cue type (No Cue, Valid, Invalid, and Center Cue), absolute Multisensory Response Enhancement (MRE) was calculated by subtracting the audiovisual median response time from the fastest unimodal median response time from the same cue type for each participant. The resulting values reflect the absolute amount of speed up or slowing down in milliseconds in the audiovisual condition compared to the fastest unimodal median response time for each cue type and for each participant.

In order to test whether possible MRE in each condition could be explained by statistical facilitation (i.e., independent processing) or by MSI, we used the individual cumulative distributive functions (CDFs) of each Target Modality for each cue type to calculate the race model using the equality (Raab, 1962):<sup>1</sup>

$$arrayP(RT_{Race Model} < t) = P(RT_A < t) + P(RT_V < t) - P(RT_A < t) \times P(RT_V < t)$$

The race model provides the probability (P) of a RT that is less than a given time  $t$  in milliseconds, where  $t$  ranges from 100–1000 ms after stimulus onset. The race model is based on the combination of the unimodal auditory and unimodal visual CDFs. We compared the observed RTs of the audiovisual CDF of each participant of each cue type to its corresponding race model (e.g., Valid Cue audiovisual CDF vs. Valid Cue race model) at the 10th, 20th, to the 90th percentile of each CDF to test for race model violations (Miller, 1982; Ulrich, Miller, & Schöter, 2007; the resulting  $p$ -values were Bonferroni corrected, see [Statistical analysis](#)). Significant violations of the race model (i.e.,  $RT_{AV} < RT_{Race model}$ ) indicate multisensory interactions that exceed statistical facilitation.

There are multiple ways in which the race model can be compared to the audiovisual CDFs. One possibility is to compare differences in RT at each quantile, as we described above. Another way to compare the CDFs is to look at differences in probability at each RT for the full function (not just taking the values of certain quantiles; see Laurienti, Burdette, Maldjian, & Wallace, 2006, Fig. 1b, for example). Using the latter option, subtracting the race model function from the audiovisual function results in a difference function showing exactly in which RT range the race model was violated.

**Statistical analysis** Overall accuracy was very high regardless of cue type (see [Results](#) section). The accuracy for the different target modalities (A, V, AV) was compared between cue types for go-trials (No Cue, Valid, Invalid, Center Cue) and for No-Go trials (No Cue, Cued, Uncued). We performed a  $3 \times 4$  repeated-measures ANOVA with the factors Target Modality (Auditory, Visual, Audiovisual) and Cue Type (No Cue, Valid, Invalid, Center Cue) for go trials, and for No-Go trials a  $3 \times 3$

<sup>1</sup> This form of the race model equality is used in several recent studies on multisensory integration in which RTs to audiovisual stimuli are compared to RTs predicted by the race model (e.g., Molholm, et al., 2006; Stevenson, Krueger, Fister, Barnett, Nidiffer, & Wallace, 2012). The equality, however, can also be expressed as  $P(RT_{Race Model} < t) = P(RT_A < t) + P(RT_V < t)$ , leaving the second part out of the equation under the assumption that responses to signals on different channels compete for resources and may therefore be negatively correlated (see Miller, 1982). Given that both are being used in the literature, we also analyzed our data using the latter formula, which provided qualitatively the same results.

repeated-measures ANOVA with the factors Target Modality (A, V, AV) and Cue Type (No Cue, Cued, Uncued) was used to analyze accuracy.

To further explore possible differences in detection performance for the different target modalities and cue types,  $A$  was calculated as a measure of sensitivity based on the hits and false alarms in each condition (Zang & Mueller, 2005; the non-parametric measure  $A$  was used as a measure for sensitivity rather than  $d$ -prime because the mean accuracy for most participants was either 1 or 0). A  $3 \times 3$  repeated-measures ANOVA with the factors Target Modality (A, V, AV) and Cue Type (No Cue, Cued, Uncued) was used to analyze possible differences in sensitivity. A distinction was made between Cued and Uncued trials instead of maintaining the Valid, Invalid, and Center Cue grouping for this analysis, because targets that were presented at the center could only be cued (Center Cue No-go trial) or uncued (No-go trial with a cue presented from the left or the right of fixation). The Go and No-go trials had to contain the same cue conditions to be able to calculate sensitivity. Therefore, in addition to a Valid Cue Go trial condition, the Invalid and Center Cue Go trials were included in an Uncued Go trial condition.

To analyze response times, a  $3 \times 4$  repeated-measures Analysis of Variance (ANOVA) was used with the factors Target Modality (Auditory, Visual, Audiovisual) and Cue Type (Valid, Invalid, Center, and No Cue). Planned pairwise comparisons between the levels of Cue Type were performed to investigate whether cueing effects were present for each Target Modality (e.g., Valid Cue RTs < Invalid Cue RTs for visual targets). We also used pairwise comparisons to investigate multisensory facilitation for each Cue Type (e.g., RT Valid Cue AV vs. RT Valid Cue V or A).

One-sample  $t$ -tests were used to test for the presence of Multisensory Response Enhancement for each Cue Type. A repeated-measures ANOVA with the factor Cue Type (No Cue, Valid, Invalid, Center) was used to test for overall differences in MRE between the different Cue Types. Planned pairwise comparisons were used to look at the differences in MRE between pairs of Cues.

Race model violations (differences in ms for each quantile) were analyzed using paired samples  $t$ -tests for each quantile. The resulting  $p$ -values were Bonferroni corrected for the number of tests within a condition ( $N = 9$  as there were nine quantiles) using the formula:  $p_{corrected} = 1 - (1 - p)^n$  (Motulsky, 1995). The second type of race model violation was analyzed using one-sample  $t$ -tests on the difference in probability between the audiovisual and the race model CDF for each ms from 0 to 1000 ms and in each condition ( $p$ -values were not Bonferroni corrected, but only violations across more than 50 consecutive RTs were reported). If the race model is violated, this would indicate MSI.

To test for differences in the amount of race model violation between cue types the median amount of race model

violation across the nine percentile points of the CDF of each cue type of each participant was used in a repeated-measures ANOVA with the factor Cue Type (No Cue, Valid, Invalid, and Center Cue), followed by planned pairwise comparisons.

The positive area under the difference curve (i.e., the difference in probability of the AV CDF and the race model CDF for the RT range of 0 to 1000 ms) was also used to investigate differences in race model violation between cue types. In order to extract the positive area under the curve for each participant, the difference curve between the AV CDF and the race model CDF was calculated for each participant. Next, all negative probabilities (no race model violation) were set to a value of zero and only the positive area under the curve was calculated for all participants. A repeated-measures ANOVA with the factor Cue Type (No Cue, Valid, Invalid, Center) was used to test for differences in the positive area under the curve, followed by planned pairwise comparisons.

In each analysis the degrees of freedom were corrected using the Greenhouse-Geisser correction when necessary.

## Results

### Accuracy

Overall accuracy on Go trials was very high (average accuracy on Go trials across all cue types,  $M = .996$ , minimum accuracy = .980, maximum accuracy = 1.00). There was a main effect of Target Modality on Go trials [ $F(1.063, 15.949) = 6.057, p = .024, \epsilon = .532, \eta^2_{partial} = .288$ ]. Bonferroni corrected pairwise comparisons between target modalities revealed a significant difference between accuracy for Auditory ( $M = 0.99, SE = .005$ ) and Audiovisual Go trials ( $M = 1.00, SE = 0.00, t(15) = -2.043, p = .047$ ), but not between Visual ( $M = 1.00, SE = .001$ ) and Audiovisual, nor Visual and Auditory go trials (all  $p$ 's > .1). There was no main effect for Cue Type and no interaction between Target Modality and Cue Type (all  $p$ 's > .1), indicating that there was no difference in the amount of anticipations (RTs < 100 ms) and misses (no response or RTs > 1000 ms) between the different cue conditions and target modalities.

For the No-go trials, there was a main effect of Target Modality [ $F(2, 30) = 7.341, p = .003, \eta^2_{partial} = .329$ ]. Participants were significantly better in withholding their response on Visual No-go trials ( $M = .960, SE = .008$ ) as compared to Auditory No-go trials ( $M = .896, SE = .019, t(15) = -4.606, p = .001$ ). There were no significant differences between Auditory and Audiovisual No-go trials ( $M = .916, SE = .017, t(15) = -0.979, p = .761$ ) and between Visual and Audiovisual No-go trials [ $t(15) = -2.686, p = .050$ ]. There was no main effect of Cue Type and there was no interaction between Cue Type and Target Modality for accuracy on No-go trials (all  $p$ 's > .05), indicating that the number of false

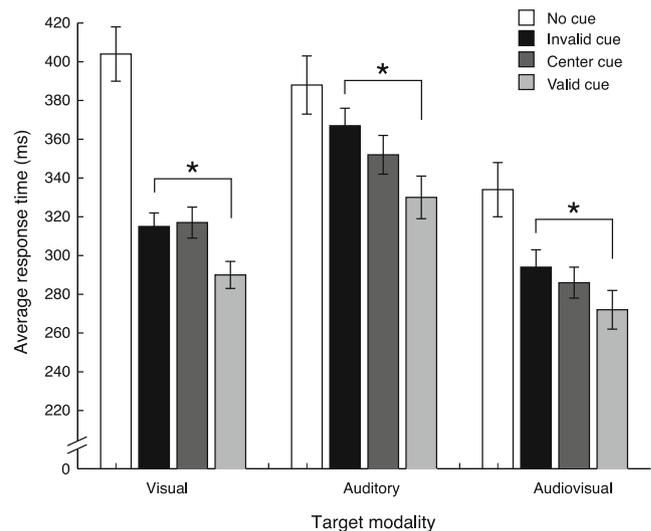
alarms did not differ between the different cue conditions and target modalities.

The analysis of sensitivity (A) revealed a main effect of Target Modality [ $F(2, 30) = 9.764, p = .001, \eta^2_{\text{partial}} = .394$ ]. Sensitivity for Visual targets was higher ( $M = .990, SE = .002$ ) compared to Auditory ( $M = .970, SE = .005; t(15) = -4.502, p = .001$ ), but comparable when compared to Audiovisual targets ( $M = .979, SE = .004, t(15) = 2.538, p = .067$ ). There was no difference in sensitivity between Auditory and Audiovisual targets [ $t(15) = -1.879, p = .221$ ]. The main effect of Cue Type [ $F(2, 30) = 1.962, p = .158, \eta^2_{\text{partial}} = .116$ ] and the interaction between Target Modality and Cue Type were not significant [ $F(4, 60) = .414, p = .798, \eta^2_{\text{partial}} = .027$ ], indicating that the sensitivity for detecting targets did not depend on the type of cue and the Target Modality.

### Response times

A significant main effect of Cue Type was found (Valid, Invalid, Center, and No Cue;  $F(1.575, 23.618) = 50.973, p < .001, \varepsilon = .525, \eta^2_{\text{partial}} = .773$ ). Figure 2 shows the mean RTs for each Target Modality for each cue type. Only the validity effects are indicated with an asterisk to keep the figure clear. Pairwise comparisons indicated that the presentation of a cue resulted in faster responses compared to the No Cue condition ( $M = 375 \text{ ms}, SE = 13$ ), regardless of whether it was a valid ( $M = 297 \text{ ms}, SE = 9, t(15) = 9.422, p < .001$ ), invalid ( $M = 324 \text{ ms}, SE = 8, t(15) = 6.083, p < .001$ ), or a central cue ( $M = 318 \text{ ms}, SE = 8, t(15) = 6.527, p < .001$ ). The speed-up in RT as a result of the presence of a cue in the second block points to the presence of a general alerting effect and an effect of a temporal warning. More importantly, however, RTs in the Valid Cue condition were significantly shorter compared to the Invalid Cue [ $t(15) = -8.251, p < .001$ ] and the Center Cue condition [ $t(15) = -4.443, p < .003$ ]. There was no significant difference in RT between the Invalid and Center Cue condition [ $t(15) = 2.027, p = .315$ ], indicating that RTs to targets following a Center Cue were much like RTs to targets following an Invalid Cue.<sup>2</sup>

Additionally, there was a main effect of Target Modality [ $F(1.390, 20.856) = 81.117, p < .001, \varepsilon = .695, \eta^2_{\text{partial}} = .844$ ]. RTs on audiovisual target trials ( $M = 297 \text{ ms}, SE = 10$ ) were significantly shorter compared to RTs on visual ( $M = 331, SE = 8, t(15) = 9.198, p < .001$ ), and auditory target trials ( $M = 359 \text{ ms}, SE = 10, t(15) = 14.685, p < .001$ ), indicative of multisensory facilitation of response times. Responses to auditory targets ( $M = 359 \text{ ms}, SE = 10$ ) were significantly slower



**Fig. 2** Mean response times to visual, auditory, and audiovisual targets for each cue type: No Cue (white), Invalid Cue (black), Center Cue (dark grey), and Valid Cue (light grey). Only the validity effects are indicated with an asterisk to maintain a clear view of the bar graph ( $p < .05$ ). See the text for more statistics on the differences between target modalities and cue types

compared to visual targets ( $M = 331 \text{ ms}, SE = 8, t(15) = 4.362, p = .003$ ). The observation that RTs to auditory targets were generally slower compared to RTs to visual targets can be explained by the fact that auditory localization is generally more difficult than visual localization (e.g., Frens & Van Opstal, 1995).

The interaction between Cue Type and Target Modality was also significant [ $F(3.588, 53.813) = 14.219, p < .001, \varepsilon = .598, \eta^2_{\text{partial}} = .487$ ]. This interaction could be explained by differences in the size of validity effects for different target modalities, by varying differences in RT between target modalities across the different cue types, or a combination of both.

To investigate the cause of the interaction, we first used pairwise comparisons to check whether cueing effects (CE; difference in RT between validly cued and invalidly cued targets) were present for each Target Modality and whether they were different in size. The difference in RT between Valid and Invalid cues was significant for auditory (mean CE = 36 ms,  $M \text{ valid} = 330 \text{ ms}, SE = 11$  vs.  $M \text{ invalid} = 366 \text{ ms}, SE = 9, t(15) = -5.398, p < .001$ ), visual (mean CE = 25 ms,  $M \text{ valid} = 290 \text{ ms}, SE = 7$  vs.  $M \text{ invalid} = 315 \text{ ms}, SE = 7, t(15) = -5.967, p < .001$ ), and audiovisual targets (mean CE = 22 ms,  $M \text{ valid} = 272 \text{ ms}, SE = 10$ , vs.  $M = 294 \text{ ms}, SE = 9, t(15) = -6.267, p < .001$ ). There were no significant differences between target modalities in the size of the validity effect (mean validity effect = 27 ms,  $SE = 3$ , all  $p$ 's  $> .22$ ). Therefore, these results do not explain the interaction, but they do indicate that the exogenous auditory cue caused an exogenous shift of attention that facilitated responses to unimodal (in a crossmodal and in an intramodal way) and bimodal targets

<sup>2</sup> The centrally presented cue is in fact an invalid cue, but presented at a smaller distance from the target location compared to the Invalid Cue condition. Still, it is difficult to say for sure whether audiovisual targets following center cues were diffusely attended or entirely unattended (given that it was an invalid cue).

that were presented at the same location as the cue. The validity effects are clearly visible in Fig. 2.

To further investigate the interaction between Cue Type and Target Modality, we compared differences between target modalities for each Cue Type using pairwise comparisons. In the No Cue condition, responses to audiovisual targets ( $M = 334$  ms,  $SE = 14$ ) were faster as compared to auditory targets [ $M = 388$  ms,  $SE = 15$ ,  $t(15) = 7.580$ ,  $p < .001$ ] as well as to visual targets ( $M = 404$  ms,  $SE = 14$ ,  $t(15) = 7.612$ ,  $p < .001$ ). The difference in RT between auditory and visual targets was not significant [ $t(15) = -1.329$ ,  $p = .495$ ].

In the Valid Cue condition, RTs on audiovisual target trials ( $M = 272$  ms,  $SE = 10$ ) were shorter compared to RTs on auditory ( $M = 330$  ms,  $SE = 11$ ,  $t(15) = 10.629$ ,  $p < .001$ ) and visual target trials ( $M = 290$  ms,  $SE = 7$ ,  $t(15) = 3.469$ ,  $p = .010$ ). Furthermore, responses to validly cued auditory targets were significantly slower compared to responses to validly cued visual targets [ $t(15) = 5.118$ ,  $p < .001$ ].

A similar pattern was observed for invalidly cued targets and centrally cued targets. In the Invalid Cue condition, RTs on audiovisual target trials ( $M = 294$  ms,  $SE = 9$ ) were shorter than on visual target trials ( $M = 315$ ,  $SE = 7$ ,  $t(15) = 4.175$ ,  $p < .002$ ) and auditory target trials ( $M = 366$ ,  $SE = 9$ ,  $t(15) = 11.194$ ,  $p < .001$ ). The difference in RT between invalidly cued visual and auditory targets was also significant [ $t(15) = 6.175$ ,  $p < .001$ ], with invalidly cued visual targets responded fastest to.

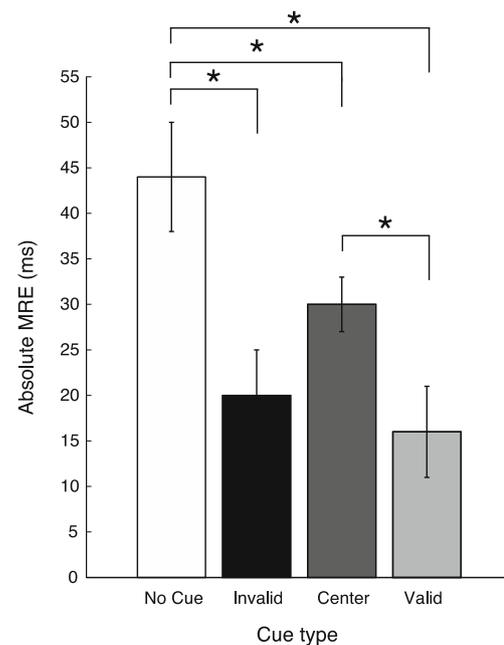
In the Center Cue condition, responses to audiovisual targets ( $M = 286$  ms,  $SE = 8$ ) were faster than to auditory ( $M = 352$  ms,  $SE = 10$ ,  $t(15) = 10.474$ ,  $p < .001$ ) and visual targets ( $M = 317$ ,  $SE = 8$ ,  $t(15) = 9.321$ ,  $p < .001$ ). Responses to visual targets were faster than to auditory targets in the Center Cue condition [ $t(15) = 5.822$ ,  $p < .001$ ].

In sum, these results indicate that whereas there was no difference in RT between auditory and visual targets in the No Cue condition, this was the case in all cued conditions, which explains the interaction.

### Multisensory response enhancement

Significant absolute MRE was observed in all cue conditions as indicated by one-sample t-tests [No Cue:  $M = 44$  ms,  $SE = 6$ ,  $t(15) = 6.937$ ,  $p < .001$ ; Valid Cue:  $M = 16$ ,  $SE = 5$ ,  $t(15) = 3.316$ ,  $p = .005$ ; Invalid Cue:  $M = 20$ ,  $SE = 5$ ,  $t(15) = 4.175$ ,  $p = .001$ ; and center cue:  $M = 30$ ,  $SE = 3$ ,  $t(15) = 10.087$ ,  $p < .001$ ]. The mean absolute MRE for each cue type is shown in Fig. 3.

To test for difference in the amount of MRE, a repeated-measures ANOVA was used. There was a main effect of Cue Type [ $F(1.919, 28.791) = 7.854$ ,  $p = .002$ ,  $\varepsilon = .640$ ,  $\eta^2_{\text{partial}} = .344$ ]. Planned pairwise comparisons indicated that MRE was significantly larger in the No Cue condition ( $M = 44$  ms,  $SE = 6$ ) compared to the Valid ( $M = 16$  ms,  $SE = 5$ ,  $t(15) = 3.658$ ,  $p = .002$ ), the Invalid ( $M = 20$  ms,  $SE = 5$ ,  $t(15) = 3.059$ ,  $p = .008$ ), and the Center Cue Condition ( $M = 30$  ms,  $SE = 3$ ,  $t(15)$



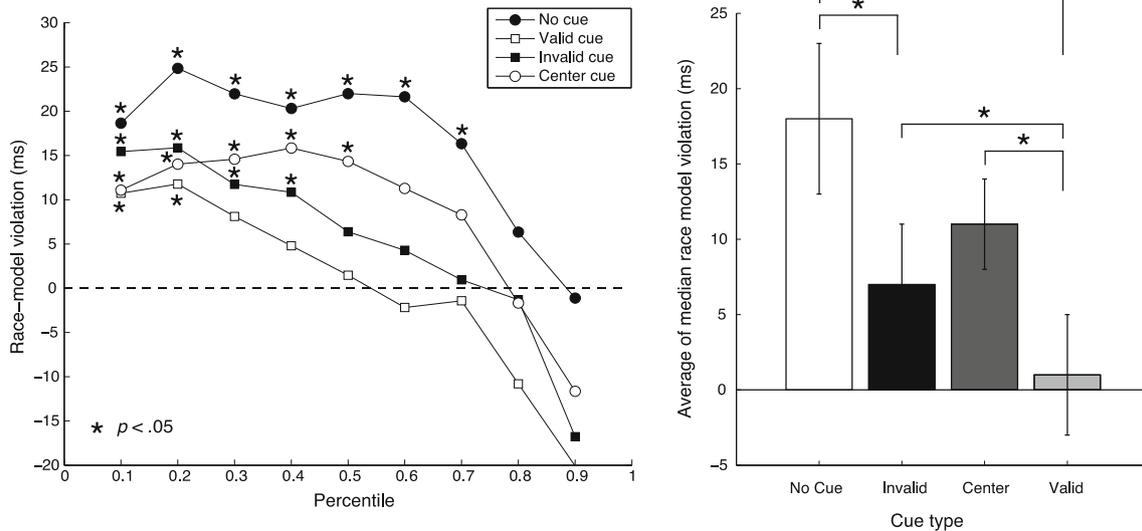
**Fig. 3** Mean absolute multisensory response enhancement (MRE) in milliseconds for each cue type: No Cue (white), Invalid Cue (black), Center Cue (dark grey), Valid Cue (light grey). Significant differences are indicated with an asterisk ( $p < .05$ )

$= 2.158$ ,  $p = .048$ ). The amount of MRE was larger in the Center compared to the Valid Cue condition [ $t(15) = -2.307$ ,  $p = .036$ ]. The differences in MRE between the Valid and the Invalid condition [ $t(15) = -1.508$ ,  $p = .152$ ] and the Invalid and Center condition [ $t(15) = -1.780$ ,  $p = .095$ ] were, however, not significant.

Center Cues are different from the No Cue and the other Cue conditions in that they were presented at the No-go location. Therefore, response inhibition to targets presented at the central location may have partly canceled the attention attracting effect of the Center Cue, resulting in a pattern of multisensory enhancement that is much more like the No Cue condition as compared to the other Cue Types (i.e., Valid and Invalid).

### Race model violation

To investigate whether the race model could explain the speed-up in the audiovisual condition in each cue condition, we compared the audiovisual CDF and the race model CDF for each cue type for each quantile bin of each CDF. The results revealed significant race model violations for all cue types. Figure 4 (left panel) shows the average amount of race model violation in each of the cue conditions for each of the quantiles. A significant race model violation was observed for the 10th to the 70th percentile in the No Cue condition ( $p < .05$ ), for the 10th to the 50th quantile in the Center Cue



**Fig. 4** Left panel: Average race model violation for each Cue condition. No Cue (filled black circle), Center Cue (open circle), Valid Cue (open square), and Invalid Cue (filled black square). Right panel: Average of the

median race model violation across all percentiles for each cue condition. Significant differences are indicated with an asterisk ( $p < .05$ )

condition ( $p < .05$ ), for the 10th to the 40th percentile in the Invalid Cue condition ( $p < .05$ ), and for the 10th and the 20th percentile in the Valid Cue condition ( $p < .05$ ). The race model could only be rejected for the fastest RTs in the Valid Cue condition, whereas this was true for a broader range of RTs in the Invalid, Center, and No Cue condition.

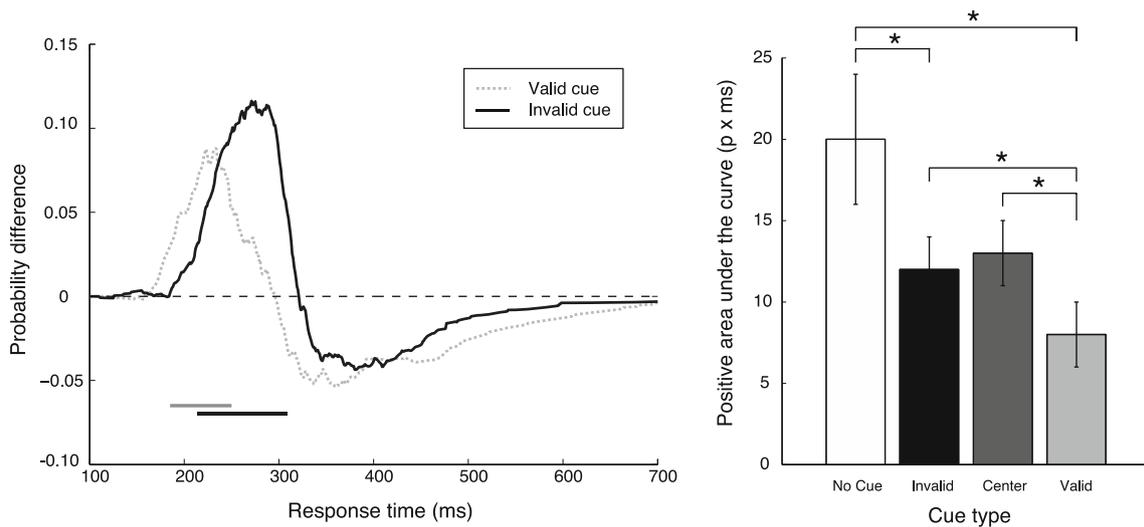
We observed a significant main effect of Cue Type on the average of the median race model violation [ $F(3,45)=5.601$ ,  $p = .002$ ,  $\eta^2_{\text{partial}} = .272$ ]. The average of the median amount of race model violation across quantiles in each Cue condition is shown in the right panel of Fig. 4. Planned pairwise comparisons revealed that the amount of race model violation in the No Cue condition ( $M=18$  ms,  $SE=5$ ) was significantly larger compared to the Invalid Cue ( $M=7$  ms,  $SE=4$ ,  $t(15)=2.247$ ,  $p = .040$ ), and the Valid Cue condition ( $M=1$  ms,  $SE=4$ ,  $t(15)=3.466$ ,  $p = .003$ ), but not compared to the Center Cue condition ( $M=11$  ms,  $SE=3$ ,  $t(15)=1.447$ ,  $p = .168$ ). More importantly, although there was no difference in absolute MRE between the Valid and the Invalid Cue condition, there was a difference in race model violation between the Valid and Invalid Cue condition [ $t(15) = -2.634$ ,  $p = .019$ ]. The difference between the Valid and Center Cue condition was also significant [ $t(15) = -2.314$ ,  $p = .035$ ], but there was no difference in the average of the median amount of race model violation between the Invalid and Center Cue condition [ $t(15) = -1.122$ ,  $p = .279$ ].

The audiovisual CDF and the race model CDF can be compared for comparable quantiles resulting in a difference in ms between two CDFs (as shown in Fig. 4), but it can also be compared in terms of differences in probability for comparable RTs (see Fig. 5; see also Laurienti, Burdette, Maldjian, & Wallace, 2006, for example). The finding that the race model

was violated at a larger range of quantiles in the Invalid Cue condition as compared to the Valid Cue condition was also supported by the range of RTs in which the race model was violated in terms of probability enhancement (see Fig. 5, left panel). In the No Cue condition the RT range in which the race model was positively violated was 142 ms long (from 244 to 385 ms). In the Valid Cue condition this window was 66 ms long (185–250 ms), which was smaller compared to the Invalid Cue condition in which this window was 96 ms long and there was a shift to somewhat slower response times (214–309 ms). The size of the window of RTs in which the race model was violated in the Center Cue condition was similar to that in the No Cue condition, but shifted to faster RTs (window =135 ms, from 192–326 ms).

The positive area under the curve was also compared between the different Cue conditions (see Fig. 5, right panel). The repeated-measures ANOVA revealed a main effect of Cue Type [ $F(1.960, 29.396)=5.242$ ,  $p = .012$ ,  $\eta^2_{\text{partial}} = .259$ ,  $\epsilon = .653$ ]. In line with the other measures of race model violation, the positive area under the curve was significantly larger in the Invalid Cue condition ( $M=12$  ms,  $SE=2.3$ ) compared to the Valid Cue condition ( $M=8$  ms,  $SE=2$ ,  $t(15)=-2.356$ ,  $p = .033$ ). The average positive area under the curve was significantly larger in the Center Cue ( $M=13$  ms,  $SE=2$ ) as compared to the Valid Cue condition [ $t(15) = -2.173$ ,  $p = .046$ ], but not compared to the Invalid Cue condition [ $t(15) = -.635$ ,  $p = .535$ ] and the No Cue condition ( $M=20$  ms,  $SE=4$ ,  $t(15)=1.917$ ,  $p = .074$ ). The positive area under the curve in the No Cue condition was significantly different from the Valid [ $t(15)=3.337$ ,  $p = .005$ ] and the Invalid Cue condition [ $t(15)=2.138$ ,  $p = .049$ ].

Given that responses in the Valid Cue condition were generally faster as compared to the Invalid Cue condition,



**Fig. 5** Left panel: Mean probability enhancement over the race model across the full range of response times (RTs) for the Valid Cue (dashed grey line), and the Invalid Cue condition (solid black line). Significant violations are indicated with a horizontal bar in the graph below the x-axis indicating the RT range in which the probabilities of consecutive time

points are larger than predicted by the race model (*p*-values were uncorrected for multiple tests but had to be consecutive). Right panel: Average positive area under the curve in each of the cue conditions. Significant differences are indicated with an asterisk (*p* < .05)

one might argue that race model violations are generally smaller in the Valid Cue condition as there may be less room for improvement in terms of RTs. In order to explore this possibility, correlations between RTs to audiovisual targets and the amount of race model violation in the Valid and Invalid Cue condition were calculated. There was a significant negative correlation between RTs and the median race model violation in both the Valid (Pearson  $r = -.829, p < .001$ ) and the Invalid Cue condition (Pearson  $r = -.821, p < .001$ ). These correlations indicate that the faster responses to audiovisual targets were, the larger the race model violation was at both the validly and the invalidly cued target locations, which is in contrast with the explanation that race model violations decrease with increasing RTs to audiovisual targets. Negative correlations were also observed between unimodal RTs and race model violation in the Valid and Invalid Cue condition (all correlations were negative, but not all correlations were significant, see Table 1). Furthermore, there was no significant difference in absolute MRE between the Valid and Invalid Cue condition, indicating that the differences in race model violation were not the result of differences in the amount of absolute MRE that was possible in the Valid and Invalid Cue condition.

*Modality switch effects*

Several different factors may contribute to the observed amount of race model violation, indicating that one must be careful with interpreting race model violations as a pure measure of multisensory integration. One of the effects that can contribute to the observed amount of race model violation is the modality Switch effect (MSE; Gondan, Lange, Rösler,

& Röder, 2004; Otto & Mamassian, 2012, Spence, Nicholls, & Driver, 2001). In a typical redundant target effect paradigm (e.g., Miller, 1986), the modality of the target is randomized across trials. This randomization results, however, in differences in RTs between trials in which a Target Modality switch occurred (relative to the previous trial) as compared to No Switch trials. Modality switches have been shown to contribute to the amount of race model violation that was observed (e.g., Gondan, Lange, Rösler, & Röder, 2004; Otto and Mamassian, 2012). As the race model is based on the distribution of RTs for unimodal auditory and visual targets, building a race model using the CDFs of faster RTs (e.g., No Switch trials) will result in a faster RT prediction of the race model as compared to slower RTs (e.g., Switch trials). In order to investigate whether the MSE partly explained the amount of race model violation in our paradigm, we also analyzed the MSE for unimodal auditory visual target trials using a N-1

**Table 1** Pearson correlations between response times to unimodal targets and the corresponding median amount of race model violation for the Valid and Invalid Cue conditions

	Median race model violation Valid Cue	Median race model violation Invalid Cue
RT Audiovisual Valid	-.829*	-.729*
RT Audiovisual Invalid	-.760*	-.821*
RT Visual Valid	-.475	-.380
RT Visual Invalid	-.452	-.478
RT Auditory Valid	-.732*	-.687*
RT Auditory Invalid	-.813*	-.659*

\* *p* < .01

trial history analysis with respect to Target Modality. Switch trials were defined as trials in which a unimodal auditory or a unimodal visual target was preceded by a unimodal target of the different modality on the previous trial (e.g., a unimodal visual target on the current trial, and a unimodal auditory target on the previous trial). All other trials were considered No Switch trials (although AV to V and AV to A trials contained both a Switch and a No Switch). The number of Switch trials in the current experiment was, however, rather low, and smaller than the number of unimodal No Switch trials as indicated by a main effect of Switch Type [ $F(1, 15) = 541.710, p < .001, \eta^2_{\text{partial}} = .973$ , mean number of Switch trials = 8,  $SE = .258$ , mean number of No Switch trials = 17,  $SE = .260$ , see Fig. 6, left panel]. There was no main effect of Target Modality [ $F(1, 15) = .273, p = .609, \eta^2_{\text{partial}} = .018$ ], but there was a significant interaction between Target Modality and Switch Type [ $F(1, 15) = 11.739, p = .004, \eta^2_{\text{partial}} = .439$ ]. Although not significant after correction, this interaction seemed to be driven by a slightly large number of auditory No Switch trials as compared to visual No Switch trials [ $t(15) = 2.766, p = .055$ ]. There was no significant difference between the number auditory and visual Switch trials [ $t(15) = -2.192, p = .168$ ].

The MSE was not analyzed for the Cue block because an auditory cue was always presented between consecutive targets. Thus, there was never a modality switch between the auditory exogenous cue and the auditory target (intramodal cueing), whereas there was always a modality switch for visual target trials (crossmodal cueing). In trials in which an audiovisual target was present, both a Switch and a No Switch occurred (i.e., both intra- and crossmodal cueing). We therefore decided that a trial history analysis would not make sense in the Cue block, because modality switches were fixed for each Target Modality and the same for each cue condition.

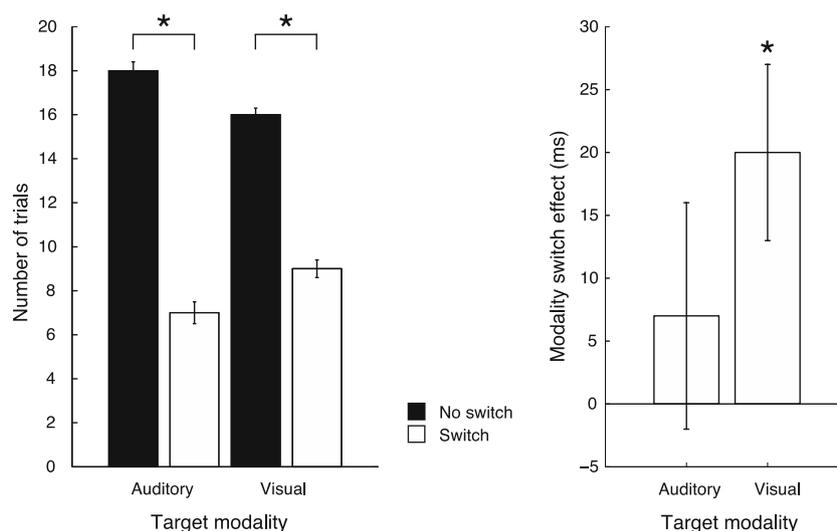
The repeated-measures ANOVA with the factors Target Modality (A, V) and Switch Type (Switch, No Switch) revealed no effects of Target Modality [ $F(1, 15) = .000, p = 1.000, \eta^2_{\text{partial}} = .000$ ], Switch Type [ $F(1, 15) = 4.130, p = .060, \eta^2_{\text{partial}} = .216$ ], nor an interaction between Target Modality and Switch Type [ $F(1, 15) = 1.689, p = .213, \eta^2_{\text{partial}} = .101$ , see Fig. 6, right panel]. We also directly analyzed the size of the MSE (the difference in RT between Switch and No Switch trials) for unimodal auditory and unimodal visual targets using one-sample t-tests, and observed an MSE for unimodal visual targets [ $t(15) = 2.794, p = .014$ ], but not for auditory targets [ $t(15) = .703, p = .493$ ].

The race model was calculated separately using unimodal Switch and unimodal No Switch trials to further explore the contribution of MSE to race model violations in the No Cue condition. The audiovisual CDF was then compared to both the Switch and the No Switch race models. For two participants there were not enough unimodal auditory Switch trials ( $N = 4$  for both participants) to calculate the race model so the race

model violation analysis was performed on data from the remaining fourteen participants. On average the No Switch race model was a bit faster as compared to the Switch race model. There were fewer percentiles at which the race model was significantly violated during No Switch trials (significant violations at the 10th and 70th percentile) as compared to Switch trials (significant violations at the 10th, 20th, 30th, and 70th percentile, see Fig. 7, left panel). The average of the median amount of race model violation for the Switch and the No Switch condition was still significant as indicated by one sample t-tests (No Switch:  $M = 16$  ms,  $SE = 6$  ms,  $t(13) = 2.238, p = .043$ ; Switch:  $M = 26$  ms,  $SE = 7$  ms,  $t(13) = 3.659, p = .003$ ). There was no significant difference in the average of the median amount of race model violation between the Switch and the No Switch condition [ $t(13) = -1.978, p = .069$ , see Fig. 7, right panel]. Thus, although MSE did seem to contribute a little to the amount of race model violation, the observed difference was not significant. These results must be interpreted with care because the number of No Switch and especially the number of Switch trials was very low (see Fig. 6).

## Discussion

The goal of the present study was to investigate whether exogenous spatial attention was able to change the outcome of multisensory integration. In order to do so, valid, invalid, and center exogenous auditory cues were presented before unimodal and bimodal audiovisual targets. The results of Experiment 1 revealed significantly larger race model violations at exogenously unattended compared to exogenously attended locations. These results suggest that multisensory integration is decreased at exogenously attended locations as compared to exogenously unattended locations. A simple explanation for this effect could be that response times cannot be much faster at attended locations and therefore results in less benefit of multisensory stimulation. There are two arguments that can be made against this explanation. First, there were significant negative correlations between response times to unimodal auditory/visual/audiovisual targets and the corresponding amount of race model violations in both the Valid and the Invalid Cue condition. The amount of race model violation increased as responses became faster, both in the Valid and in the Invalid Cue condition. This indicates that the amount of race model violation did not systematically decrease as the absolute RTs decreased (i.e., faster responses did not result in less race model violation). Furthermore, the absolute amount of MRE did not significantly differ between the validly and the invalidly cued targets, indicating that in terms of absolute speed up in the multisensory condition compared to the fastest unimodal condition, there was no difference between attended and unattended locations. Crucially, most of the multisensory speed up at exogenously attended locations could, however, be explained by statistical facilitation, whereas this was not the case at exogenously unattended locations.



**Fig. 6** Left panel: Mean number of Switch and No Switch trials for unimodal auditory and visual targets in the No Cue condition. An asterisk indicates a significant difference between conditions ( $p < .05$ ). Right

panel: The modality switch effect for auditory and visual targets. The asterisk indicates a significant difference from zero ( $p < .05$ )

One could argue that the difference in sensitivity between visual targets and auditory targets makes the interpretation of the results difficult. However, audiovisual targets always consisted of the same unimodal components and the critical manipulation of interest was a possible modulation of MRE by exogenous spatial attention. There were no differences in sensitivity between cue conditions and no interaction between cue condition and Target Modality. Therefore, we do not think that the observed differences in sensitivity are problematic because these differences were equal for the Valid and the Invalid Cue condition.

Race model violations can sometimes be explained by taking trial-history effects into account. In particular modality switch effects can contribute to the amount of race model violation (e.g., Otto & Mamassian, 2012). As others have shown, this is not always the case (Gondan, Lange, Rösler, & Röder, 2004). In the No Cue condition of Experiment 1, there was a slight decrease in the amount of race model violation, which could be explained by MSE. However, this decrease was not significant and the race model was still violated when the AV CDF was compared to a race model that consisted of only unimodal No Switch trials. Trial history effects were not analyzed in the cued conditions as the exogenous auditory cue always caused a modality switch for visual targets within a trial, and never for auditory targets within a trial. For audiovisual targets, both a modality Switch and a No Modality Switch occurred after the presentation of the cue.

Additionally, we observed spatial cueing effects for all target modalities. If modality switch effects contributed more to the speed up of responses to different targets than exogenous spatial attention, then validly cued unimodal auditory targets should be responded to faster as compared to unimodal visual targets, because auditory processing benefits both from within modality priming and exogenous spatial attention. This cannot

be concluded from the data, as responses to visual targets were always faster compared to responses to auditory targets in the cued conditions despite the modality switch that was always present within a trial for visual targets. Importantly, there was no difference in RTs for unimodal visual and auditory targets in the No Cue condition. We therefore argue that the current observation of larger race model violations for invalidly cued targets as compared to validly cued targets reflects differences in multisensory integration between exogenously attended versus unattended target locations.

The amount of multisensory integration has been shown to differ between localization and detection paradigms (e.g., Hecht, Reiner, & Karni, 2008). Whereas in localization paradigms participants have to process the spatial location of targets before responding, spatial information is task-irrelevant in detections paradigms. The results from previous studies suggest that multisensory integration of simple lights and sounds is especially helpful when spatial orienting is task-relevant (e.g., Hecht, Reiner, & Karni, 2008). Exogenous spatial attention may therefore have a larger influence on multisensory integration in tasks in which spatial localization is important. When spatial localization is task-irrelevant, the influence of exogenous spatial attention on multisensory integration may decrease substantially as the task can be completed without relying on spatial information. In order to test possible task dependencies of the effect of exogenous spatial attention on multisensory integration, we conducted another experiment in which a Go/No-go detection task was used. Although the task was still to detect auditory, visual, and audiovisual targets using one button, this time targets were only presented from the left and the right side of the fixation cross but not at the central location. No-go (catch) trials did not consist of the presentation of the target at the fixation

cross, but of the absence of a target. This way, participants did not have to localize the target before deciding to respond to the presence of a target in Experiment 2, which was the case in Experiment 1.

## Experiment 2

### Materials and methods

#### Participants

Twenty-four participants were tested in Experiment 2 (six male, 18 female, mean age = 25.5 years,  $SD = 3.5$ ). All participants had normal or corrected-to-normal visual acuity and did not report any hearing problems. The participants received either course credits or a monetary reward for their participation. The experiment was conducted in accordance with the Declaration of Helsinki, and participants signed informed consent before the start of the experiment.

#### Apparatus

Visual and auditory stimulus presentation was controlled through a custom-built audiovisual stimulus generator<sup>3</sup> that was connected to a PC running MATLAB. Visual stimuli consisted of tri-colored Light Emitting Diodes (LEDs, Forge Europa, bulb size: 5 mm, viewing angle: 65°, LED colour: red, green, and blue) of which the intensity and color could be adjusted independently. Auditory stimuli were presented over loudspeakers (e-audio black 4" Full Range Mini Box Speaker, dimensions: 120 × 120 × 132 mm, frequency response: 80–20,000 Hz). With the use of the audiovisual stimulus generator, stimuli were presented with high temporal precision (i.e., an accuracy of 1 ms in terms of their onset and offset during a trial, timing was confirmed using an oscilloscope). Participants were instructed to place their chin in a chin-rest to keep the distance between the stimuli and the participants the same. Three loudspeakers and three LEDs were used to present the stimuli at a distance of 64 cm from the participant. Each LED was attached to the center of a speaker ensuring precise spatial alignment. The left and the right speakers with LEDs were placed at 33° to the left and the right of the central speaker and LED.

#### Stimuli, task, and procedure

As in Experiment 1, participants had to detect auditory, visual, and audiovisual targets. This time, targets were presented only at the left and the right location, and not at the central location.

<sup>3</sup> Schemes of the device, a list of its components, and a compatible programmed chip are available upon request.

Whereas Experiment 1 was a Go/No-go localization detection task, Experiment 2 was a simple detection task. Exogenous auditory cues were present in each trial and were presented from the left or right location. Targets could be presented at the same or a different location (i.e., from the opposite side of the central fixation location) as the cue. Thus, targets could be either validly or invalidly cued. Each trial started with the onset of a central blue LED (266 cd/m<sup>2</sup>) with a random duration between 750 and 1250 ms. After the fixation offset, a 100-ms exogenous auditory cue was presented from the left or right location. The cue was a 75-ms 800 Hz pure tone [~66 dB(A)]. The target consisted of the illumination of a green LED (2072 cd/m<sup>2</sup>) for 100 ms, the presentation of a white noise sound [~66 dB(A)] for 100 ms, or the combination of the green LED and the white noise sound (SOA = 0 ms; Target Modality and target location were randomized). The cue-target stimulus onset asynchrony varied randomly between 200 and 250 ms and the response window was set to 2000 ms after target onset after which the next trial started. There were 80 auditory targets (40 validly cued, 40 invalidly cued), 80 visual targets (40 validly cued, 40 invalidly cued), and 80 audiovisual targets (40 validly cued, 40 invalidly cued). Sixty catch trials (20 %) were implemented in which no target was presented after the presentation of the cue to ensure that participants were paying attention to the task and responded only to the targets. The participants were instructed to press a button on a custom-made response box as quickly as possible after the detection of the green target LED, the white noise sound, or their combination, and to withhold their response when no target appeared after the auditory cue.

#### Data preprocessing

The same data preprocessing steps as in Experiment 1 were used.

#### Statistical analyses

The same analyses as in Experiment 1 were used with the exception that the factor Cue Type contained only two levels in Experiment 2 (Valid, Invalid).

### Results

#### Accuracy

Overall, accuracy was high on Go ( $M = .974$ ,  $SE = .007$ ) and No-go trials ( $M = .916$ ,  $SE = .100$ ).

The repeated-measures ANOVA of accuracy on Go trials revealed a main effect of Cue Type [ $F(1, 23) = 5.793$ ,  $p = .025$ ,  $\eta^2_{\text{partial}} = .201$ ], but no main effect of Target Modality [ $F(1.585, 35.466) = 1.077$ ,  $p = .338$ ,  $\varepsilon = .793$ ,  $\eta^2_{\text{partial}} = .045$ ] and no interaction between Target Modality and Cue Type [ $F(2, 46) = 1.103$ ,  $p = .341$ ,  $\eta^2_{\text{partial}} = .046$ ]. Accuracy on

Go trials in the Invalid Cue condition was generally higher ( $M = .978, SE = .007$ ) as compared to the Valid Cue condition ( $M = .969, SE = .007$ ).

The analysis of sensitivity also indicated a main effect of Cue Type [ $F(1, 23) = 5.121, p = .033, \eta^2_{partial} = .182$ ]. There was no main effect of Target Modality [ $F(1.545, 35.537) = 1.201, p = .303, \eta^2_{partial} = .050$ ] and no interaction between Target Modality and Cue Type [ $F(2, 46) = 1.355, p = .268, \eta^2_{partial} = .056$ ]. The sensitivity was higher in the Invalid Cue condition (mean  $A = .971, SE = .005$ ) as compared to the Valid Cue condition (mean  $A = .968, SE = .005$ ), but the overall accuracy was so high that we want to be careful in interpreting this difference and do not want to draw strong conclusions from this result.

*Response times*

A main effect of Target Modality was observed [ $F(1.332, 30.627) = 60.376, p < .001, \varepsilon = .666, \eta^2_{partial} = .724$ ]. RTs for audiovisual targets ( $M = 268 \text{ ms}, SE = 11 \text{ ms}$ ) were significantly shorter as compared to auditory ( $M = 333 \text{ ms}, SE = 15 \text{ ms}, t(23) = 9.531, p < .001$ ) and visual targets ( $M = 329 \text{ ms}, SE = 10 \text{ ms}, t(23) = 15.347, p < .001$ ). RTs for unimodal auditory and unimodal visual targets did not differ [ $t(23) = .454, p = .654$ ].

The main effect of Cue Type was also significant [ $F(1, 23) = 42.162, p < .001, \eta^2_{partial} = .647$ ], indicating that responses to validly cued targets were generally faster ( $M = 303 \text{ ms}, SE = 12 \text{ ms}$ ) as compared to invalidly cued targets ( $M = 317 \text{ ms}, SE = 12 \text{ ms}$ , see Fig. 8, left panel). There was no interaction between Target Modality and Cue Type [ $F(2, 46) = 2.349, p = .107, \eta^2_{partial} = .093$ ].

A closer inspection of the validity effect (VE) for each Target Modality using one-sample t-tests revealed that there was a significant VE for visual ( $M = 24 \text{ ms}, SE = 5 \text{ ms}, t(23) = -5.020, p < .001$ ) and audiovisual targets ( $M = 12 \text{ ms}, SE = 3 \text{ ms}, t(23) = -3.537, p = .002$ ), but not for auditory targets ( $M = 8 \text{ ms}, SE = 6 \text{ ms}, t(23) = -1.311, p = .203$ ).

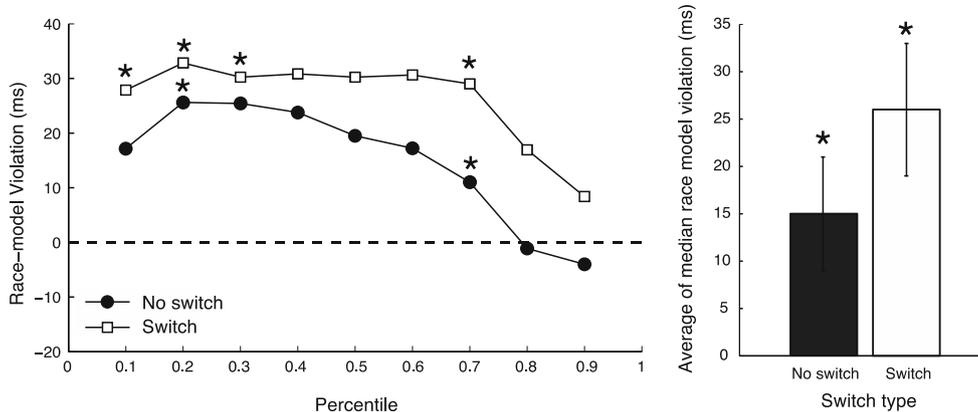
*Multisensory response enhancement*

One-sample tests revealed that there was significant absolute MRE in both the valid (mean MRE = 40 ms, SE = 5 ms) and the invalid condition (mean MRE = 52 ms, SE = 5 ms, all  $p < .001$ , see Fig. 8, right panel). The amount of absolute MRE was not significantly different between the Valid and the Invalid Cue condition (but close to significance,  $t(23) = -2.019, p = .055$ ).

*Race model violation*

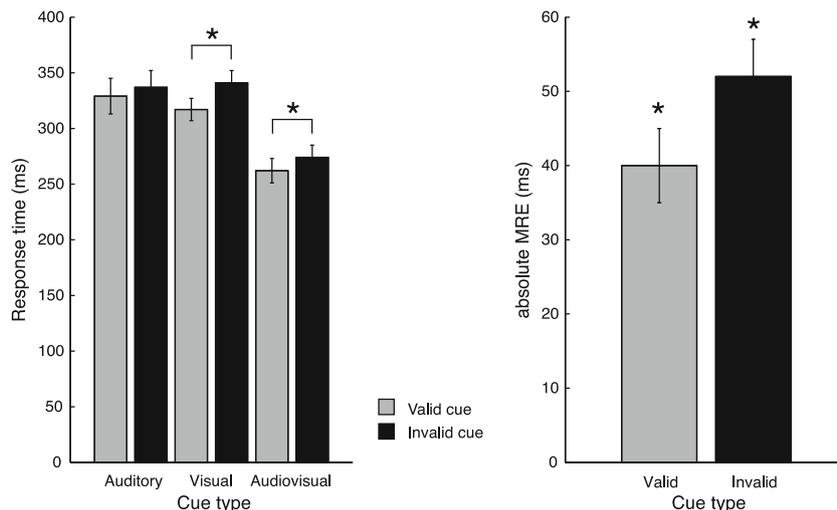
Significant race model violations were observed only in the Invalid Cue condition, not in the Valid Cue condition. The race model was significantly violated at the 40th and 50th percentiles ( $p < .05$ , 2-tailed, Bonferroni corrected, see Fig. 9, left panel). There was no difference in the average of the median amount of race model violation between the Valid ( $M = 6 \text{ ms}, SE = 4$ ) and the Invalid Cue conditions ( $M = 7 \text{ ms}, SE = 4, t(23) = -.294, p = .771$ , see Fig. 9, right panel).

A more detailed look at the range of RTs in which the race model was violated at consecutive RTs is provided in Fig. 10 (left panel). This was the case in the 193–323 ms range (window width = 131 ms) in the Valid Cue condition. As in Experiment 1, this range was larger and shifted to later RTs in the Invalid Cue condition in which consecutive significant race model violations were observed in the RT range 207–437 ms (window width = 231 ms). The RT range in which consecutive race model violations occurred was 100 ms larger in the Invalid Cue condition as compared to the Valid Cue condition. The average positive area under the curve was significantly different from the race model both in the Valid ( $M = 19, SE = 3, t(23) = 7.047, p < .001$ ) and the Invalid ( $M = 21, SE = 15, t(23) = 7.792, p < .001$ ) Cue condition, as shown by one-sample t-tests. There was no difference in the average positive area under the curve between the Valid and Invalid



**Fig. 7** Left panel: The average race model violation in the Switch and No-Switch trials in the No Cue condition of Experiment 1. Significant differences are indicated with an asterisk ( $p < .05$ , Bonferroni corrected).

Right panel: The average of the median race model violation for Switch and No-Switch trials. Asterisks indicate a significant difference from zero ( $p < .05$ )



**Fig. 8** Left panel: The average of median response times for each Target Modality and Cue Type. Significant validity effects are indicated with an asterisk ( $p < .05$ ). Right panel: The absolute amount of MRE in the Valid

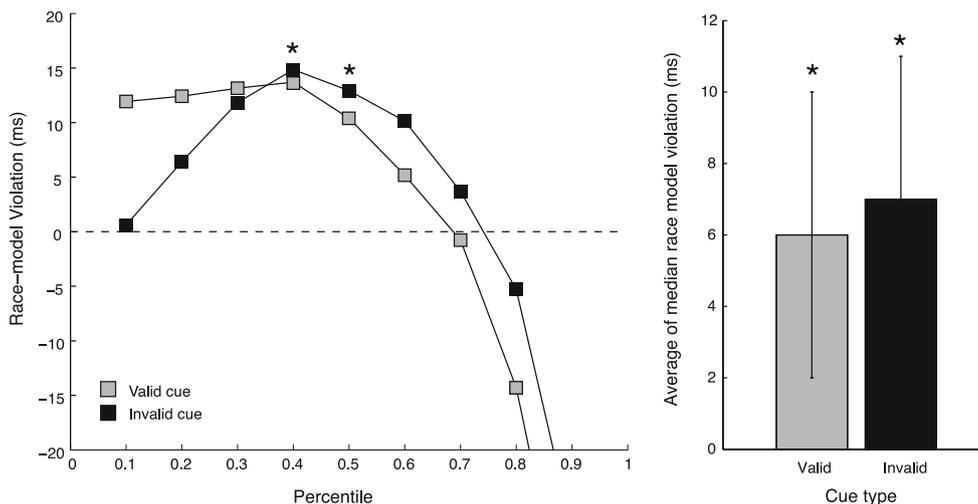
and Invalid Cue conditions. Significant absolute MRE is indicated with an asterisk ( $p < .05$ )

Cue condition [ $t(23) = -.368, p = .716$ , see Fig. 10, right panel].

Discussion

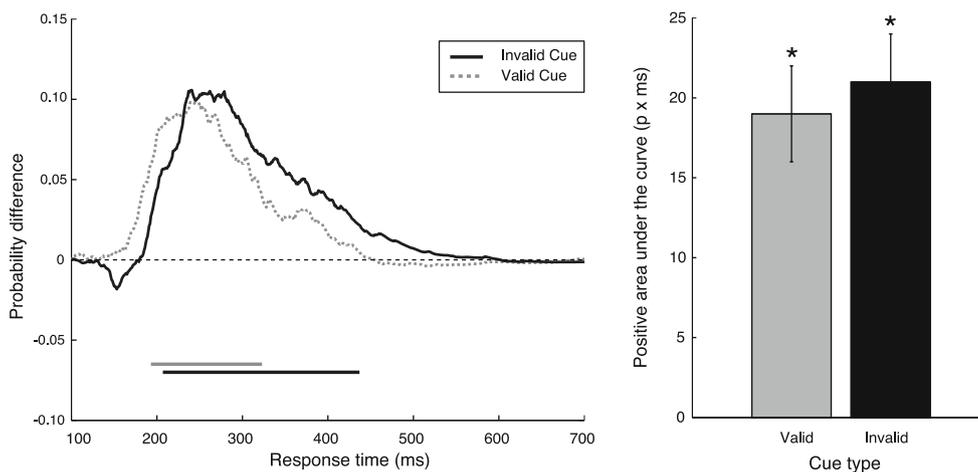
In Experiment 2 we investigated the influence of exogenous spatial attention on multisensory integration in a Go/No-go detection paradigm in order to investigate whether the interaction between exogenous spatial attention and MSI depended on spatial localization. The observed effect of exogenous spatial attention on MSI was not as pronounced in Experiment 2 as in Experiment 1. There were differences in the range in which the race model was violated, but this was not reflected in all measures. The larger amount of MSI in the

Invalid Cue condition was mainly visible in the absence of any race model violation in the Valid Cue condition (in terms of the absolute difference in ms between the race model CDF and the audiovisual CDF for different quantiles) and the broader range of RTs in which the race model was violated in terms of probability enhancement in the Invalid Cue condition. The RT range in which the race model was violated was 100 ms larger for targets appearing at exogenously unattended locations as compared to exogenously attended locations. In sum, although the spatial location of stimuli was task-irrelevant in Experiment 2, we still observed some effects of the exogenous cue on MSI. These results could be taken to suggest that the interaction between exogenous spatial attention and MSI is affected by whether spatial localization is task-relevant. This



**Fig. 9** Left panel: The average race model violations for the Valid Cue (filled grey squares) and the Invalid Cue (filled black squares) conditions. Right panel: Average of the median race model violation across all

quantiles for the Valid and Invalid Cue conditions. Significant race model violations are indicated with an asterisk ( $p < .05$ )



**Fig. 10** Left panel: Mean probability enhancement over the race model across the full range of response times (RTs) for the Valid Cue (solid black line), and the Invalid Cue (dashed grey line) conditions. Significant violations are indicated with a bar in the graph depicting the RT range at which the probabilities of consecutive time points are larger than

predicted by the race model ( $p$ -values were uncorrected for multiple tests but had to be consecutive). Right panel: The average positive area under the curve in Valid and Invalid Cue condition. Significant differences are indicated with an asterisk ( $p < .05$ )

may not be very surprising considering that exogenous spatial attention is inherently spatial and its effects on other processes may therefore be especially pronounced in spatial tasks.

## General discussion

The goal of the present study was to investigate whether exogenous spatial attention could affect multisensory integration of simple lights and sounds. This study is the first to show that exogenous spatial attention can influence MSI, given that enough time is provided for the exogenous cue to attract attention to the target location and that the cue is not part of the stimuli that need to be integrated. Although the absolute amount of MRE was the same when the target location was attended compared to when it was unattended, MSI was decreased at the attended location as compared to the unattended locations. This indicates that exogenous spatial attention speeds up the processing of multisensory stimuli (as indicated by cueing effects for unimodal and multisensory targets), but also decreases the amount of MSI. Besides the effects of exogenous spatial attention on MSI, the exogenous cue also caused an alerting effect and acted as a temporal warning, which contributed to a general decrease in MSI both at attended and unattended locations as indicated by a decrease in race model violation in the cued conditions as compared to the condition in which no cue was present. A temporal warning can increase the expectation of the appearance of a target and decrease MSI compared to a situation in which there is no temporal warning, or, in other words, a lower expectation of an upcoming target (e.g., Reches, Netser, and Gutfreund, 2010).

Several frameworks of multisensory integration have been proposed to account for the interactions (or lack of an interaction) between attention and multisensory integration (Koelewijn et al., 2010). Whereas the ‘early integration’ framework states that MSI occurs at an early pre-attentive stage, the ‘late integration’ framework states that integration takes place a late stage, leaving room for attention to affect the unimodal sensory input. Additionally, the parallel integration framework (Calvert & Thesen, 2004) states that MSI can occur at both ‘early’ and ‘late’ stages of sensory processing depending on spatial, temporal, and/or content properties of the stimuli. Several studies have provided examples of each of these stages and together provide support for the parallel integration framework (see Koelewijn et al., 2010, for a review). By presenting the exogenous spatial cue a little ahead in time of the multisensory target, exogenous spatial attention was allowed time to shift to the location of the target and decrease MSI. If we assume that the integration of spatially and temporally aligned simple lights and sounds is indeed the result of ‘early’ MSI, then we should perhaps reconsider the idea that ‘early’ integration always occurs in a pre-attentive way (as suggested by Koelewijn et al., 2010).

The effects of *exogenous* attention on MSI that we observed are opposite to those found in most studies in which the influence of *endogenous* attention on MSI was investigated (e.g. Talsma & Woldorff, 2005; Fairhall & Macaluso, 2009). Whereas previous studies have provided evidence for the idea that *endogenously* attending the target location enhances MSI of simple stimuli, we observed that *exogenously* attending the location of the target decreased MSI. Our findings are in line with the results from Zou, Müller, and Shi (2012) who found that MSI was enhanced in the *endogenously unattended* half of the stimulus field (as shown by the

strength of the pip and pop effect). A key factor in whether attention is able to decrease multisensory integration at attended locations as compared to unattended locations might be the need to localize targets in space. When targets are always presented at the same location, there is no need to pay attention to multiple spatial locations. When the target location is uncertain (in both Experiment 1 and Experiment 2 of our study and the study by Zou, Müller, and Shi, 2012), MSI is decreased at the attended location as compared to the unattended location, while the opposite seems to be true for fixed target locations (e.g., Talsma & Woldorff, 2005; Fairhall & Macaluso, 2009).

The spatial uncertainty of target locations seems to be linked to the need for spatial orienting, with a higher uncertainty of the target location leading to a higher need for spatial orienting. This was also clear from the results of Experiment 2, in which targets did not have to be localized in order to perform the task, making spatial orienting task-irrelevant. As in Experiment 1, a small decrease in MSI was observed at exogenously attended locations as compared to exogenously unattended locations, but the effect was less pronounced (as seen in the lack of, for example, a significant difference in average race model violation). Even though the spatial location of cues and targets was task-irrelevant in Experiment 2, the location of the cue and the target was still uncertain (i.e., the target randomly appeared to the left or the right of the fixation cross, or did not appear), which may have increased the chances of spatial orienting and finding similar but less pronounced effects of exogenous spatial cues as in Experiment 1. The idea that some spatial orienting still occurred is supported by the observation of a cueing effect for visual and audiovisual targets, but the absence of a spatial cueing effect for auditory targets in Experiment 2 indicates that the cueing effect was less robust as in Experiment 1. If spatial uncertainty is indeed a key factor in the interaction between exogenous spatial attention and MSI, then MSI might be decreased at attended target locations because spatial orienting that is evoked by the cue and spatial orientation caused by the multisensory target (and therefore multisensory integration) are redundant (i.e., multisensory integration is not as helpful in spatial localization when attention is already at the right location). The opposite is true for exogenously unattended or invalidly cued targets.

A further possible explanation for the decrease in MSI at exogenously attended locations as compared to exogenously unattended locations could be related to the changes in perceptual sensitivity caused by exogenous cues. Exogenous spatial attention is able to increase perceptual sensitivity at attended locations as compared to unattended locations (e.g. Carrasco, Penpeci-Talgar, & Eckstein, 2000; see Carrasco, 2011, for a review). Research has shown that the effects of exogenous attention on perceptual sensitivity can be explained by a combination of response gain and contrast gain models

(Ling & Carrasco, 2006). Whereas contrast gain is effectively similar to an increase of stimulus contrast at the attended location, response gain can be interpreted as an increase in stimulus intensity at the attended location. Exogenously attending a multisensory target could thus lead to an increase in perceived intensity and contrast. The principle of inverse effectiveness states that the benefit of multisensory integration (e.g., in terms of RT) is larger for weaker stimuli (e.g., less intense) than for stronger stimuli (e.g., more intense). One could argue that because perceptual sensitivity is higher at exogenously attended locations, MSI is decreased at exogenous attended locations through inverse effectiveness (but see Leone & McCourt, 2013; Holmes, 2007, 2009).

The current results may also be interesting from a more applied point of view. Exogenously attended multisensory targets were responded to the fastest in an absolute sense, compared to unattended multisensory targets. Thus, regardless of whether or not integration occurred when the target location was attended compared to when it was unattended, participants benefited the most from the combination of an exogenous cue and a multisensory stimulus, as shown by their RTs. This may be especially relevant in designing multisensory interfaces when the focus is on the speed of responding.

As far as we know, this is the first study to report that exogenous spatial attention influences audiovisual integration. Exogenous attention was able to speed up responses to multisensory stimuli, but at the same time also decreased the amount of MSI. These findings are in contrast with the idea that ‘early’ integration cannot be affected by exogenous spatial attention. Last, there may be an important role for spatial orienting and spatial uncertainty in this interaction, as indicated by a more pronounced interaction when the location of stimuli was task-relevant compared to when the location of the stimuli was task-irrelevant.

**Acknowledgment** This research was funded by two grants from the NWO (Netherlands Organization for Scientific Research): Grant Nos. 451-09-019 (to S.V.d.S.) and 451-10-013 (to T.C.W.N.).

## References

- Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*, 839–843.
- Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research*, *183*, 399–404.
- Berger, A., Henik, A., & Rafal, R. (2005). Competition between endogenous and exogenous orienting of visual attention. *Journal of Experimental Psychology: General*, *134*(2), 207.
- Bertelson, P., Pavani, F., Ladavas, E., Vroomen, J., & de Gelder, B. (2000a). Ventriloquism in patients with unilateral visual neglect. *Neuropsychologia*, *38*, 1634–1642.

- Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000b). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, *62*(2), 321–332.
- 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. *PNAS*, *102*(51), 18751–18756.
- Calvert, G. A., & Thesen, T. (2004). Multisensory integration: Methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris*, *98*(1), 191–205.
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*, 1484–1525.
- Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial cover attention increases contrast sensitivity across CSF: Support for signal enhancement. *Vision Research*, *40*, 1203–1215.
- Driver, J., & Spence, C. (1998). Cross-modal links in spatial attention. *Philosophical Transactions Royal Society of London Series B*, *353*, 1319–1331.
- Fairhall, S. L., & Macaluso, E. (2009). Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *European Journal of Neuroscience*, *29*, 1247–1257.
- Frens, M. A., Van Opstal, A. J., & Van der Willigen, R. F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Perception & Psychophysics*, *57*(6), 802–816.
- Gondan, M., Lange, K., Rosler, F., & Roder, B. (2004). The redundant target effect is affected by modality switch costs. *Psychonomic Bulletin & Review*, *11*(2), 307–313.
- Hecht, D., Reiner, M., & Karni, A. (2008). Multisensory enhancement: Gains in choice and in simple response times. *Experimental Brain Research*, *189*(2), 133–143.
- Hirsh, I. J., & Sherrick, C. E., Jr. (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology*, *62*(5), 423.
- Holmes, N. P. (2007). The law of inverse effectiveness in neurons and behavior: Multisensory integration versus normal variability. *Neuropsychologia*, *45*, 3340–3345.
- Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: Some statistical considerations. *Brain Topography*, *21*(3–4), 168–176.
- Keetels, M., & Vroomen, J. (2005). The role of spatial disparity and hemifields in audio-visual temporal order judgments. *Experimental Brain Research*, *167*(4), 635–640.
- Koelwijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stage of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, *134*, 372–384.
- Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging*, *27*(8), 1155–1163.
- Leo, F., Bolognini, N., Passamonit, C., Stein, B. E., & Ladavas, E. (2008). Cross-modal localization in hemianopia: New insights on multisensory integration. *Brain*, *131*(3), 855–865.
- Leone, L. M., & McCourt, M. E. (2013). The roles of physical and physiological simultaneity in audiovisual multisensory facilitation. *i-Perception*, *4*(4), 213.
- Ling, S., Liu, T., & Carrasco, M. (2009). How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Research*, *49*(10), 1194–1204.
- Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by cross-modal spatial attention. *Science*, *289*(5482), 1206–1208.
- Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience*, *7*(10), 3215–3229.
- McDonald, J. J., Teder-Sälejärvi, W. A., & Ward, L. M. (2001). Multisensory integration and crossmodal attention effects in the human brain. *Science*, *292*(5523), 1791.
- Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*, 247–279.
- Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics*, *40*(5), 331–343.
- Molholm, S., Sehatpour, P., Metha, A. D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., ... Foxe, J. J. (2006). Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. *Journal of Neurophysiology*, *96*, 721–729.
- Montagna, B., Pestilli, F., & Carrasco, M. (2009). Attention trades off spatial acuity. *Vision Research*, *49*(7), 735–745.
- Motulsky, H. J. (1995). *Intuitive biostatistics*. Oxford: Oxford University Press.
- Navarra, J., Alsius, A., Soto-Faraco, S., & Spence, C. (2010). Assessing the role of attention in the audiovisual integration of speech. *Information Fusion*, *11*(1), 4–11.
- Ngo, M. K., & Spence, C. (2010). Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Attention, Perception, & Psychophysics*, *72*(6), 1654–1665.
- Otto, T. U., & Mamassian, P. (2012). Noise and correlations in parallel perceptual decision making. *Current Biology*, *22*(15), 1391–1396.
- Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*, 574–590.
- Reches, A., Netszer, S., & Gutfreund, Y. (2010). Interactions between stimulus-specific adaptation and visual auditory integration in the forebrain of the barn owl. *The Journal of Neuroscience*, *30*(20), 6991–6998.
- Santangelo, V., Ho, C., & Spence, C. (2008a). Capturing spatial attention with multisensory cues. *Psychonomic Bulletin & Review*, *15*(2), 398–403.
- Santangelo, V., Van der Lubbe, R. H., Belardinelli, M. O., & Postma, A. (2006). Spatial attention triggered by unimodal, crossmodal, and bimodal exogenous cues: A comparison of reflexive orienting mechanisms. *Experimental Brain Research*, *173*(1), 40–48.
- Santangelo, V., Van der Lubbe, R. H., Belardinelli, M. O., & Postma, A. (2008b). Multisensory integration affects ERP components elicited by exogenous cues. *Experimental Brain Research*, *185*(2), 269–277.
- Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: Evidence from the speeded classification task. *Cognition*, *92*, B13–B23.
- Spence, C. (2010). Crossmodal spatial attention. *Annals of the New York Academy of Sciences*, *1191*(1), 182–200.
- Spence, C. J., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(3), 555.
- Spence, C., & Driver, C. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, *59*(1), 1–22.
- Spence, C., & Driver, C. (2000). Attracting attention to the illusory location of a sound: Reflexive crossmodal orienting and ventriloquism. *Neuroreport*, *11*(9), 2057–2061.
- Spence, C., Nicholls, M. E. R., & Driver, J. (2001). The cost of expecting events in the wrong sensory modality. *Perception & Psychophysics*, *63*(2), 330–336.
- Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: A review. *Hearing Research*, *258*(1), 134–142.
- Stevenson, R. A., Krueger Fister, J., Barnett, Z. P., Nidiffer, A. R., & Wallace, M. T. (2012). Interactions between the spatial and temporal stimulus factors that influence multisensory integration in human performance. *Experimental Brain Research*, *219*, 121–137.
- Stevenson, R. A., & Wallace, M. T. (2013). Multisensory temporal integration: Task and stimulus dependencies. *Experimental Brain Research*, *227*(2), 249–261.
- Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*, 400–410.

- Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*, *17*, 679–690.
- Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, *17*(7), 1098–1114.
- Ulrich, R., Miller, J., & Schröter, H. (2007). Testing the race model inequality: An algorithm and computer programs. *Behavior Research Methods*, *39*(2), 291–302.
- Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1053.
- Van der Burg, E., Talsma, D., Olivers, C. N., Hickey, C., & Theeuwes, J. (2011). Early multisensory interactions affect the competition among multiple visual objects. *NeuroImage*, *55*(3), 1208–1218.
- Vroomen, J., Bertelson, P., & de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, *63*(4), 651–659.
- Vroomen, J., & Gelder, B. D. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1583.
- Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics*, *72*(4), 871–884.
- Ward, L. M., McDonald, J., & Lin, D. (2000). On asymmetries in cross-modal spatial attention orienting. *Perception & Psychophysics*, *62*(6), 1258–1264.
- Zampini, M., Guest, S., Shore, D. I., & Spence, C. (2005). Audio-visual simultaneity judgments. *Perception & Psychophysics*, *67*(3), 531–544.
- Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika*, *70*(1), 203–212.
- Zou, H., Müller, H. J., & Shi, Z. (2012). Non-spatial sounds regulate eye movements and enhance visual search. *Journal of Vision*, *12*(5), 1–18.