



# Bimodal-divided attention attenuates visually induced inhibition of return with audiovisual targets

Xiaoyu Tang<sup>1,2</sup> · Yulin Gao<sup>3</sup> · Weiping Yang<sup>4</sup> · Yanna Ren<sup>5</sup> · Jinglong Wu<sup>2,6,7</sup> · Ming Zhang<sup>8</sup> · Qiong Wu<sup>2</sup>

Received: 26 January 2017 / Accepted: 4 February 2019 / Published online: 15 February 2019  
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

Inhibition of return (IOR) refers to the slower response to a target appearing at a previously attended location in a cue–target paradigm. It has been greatly explored in the visual or auditory modality. This study investigates differences between the IOR of audiovisual targets and the IOR of visual targets under conditions of modality-specific selective attention (Experiment 1) and divided-modalities attention (Experiment 2). We employed an exogenous spatial cueing paradigm and manipulated the modalities of targets, including visual, auditory, or audiovisual modalities. The participants were asked to detect targets in visual modality or both visual and auditory modalities, which were presented on the same (cued) or opposite (uncued) side as the preceding visual peripheral cues. In Experiment 1, we found the comparable IOR with visual and audiovisual targets when participants were asked to selectively focus on visual modality. In Experiment 2, however, there was a smaller magnitude of IOR with audiovisual targets as compared with visual targets when paying attention to both visual and auditory modalities. We also observed a reduced multisensory response enhancement effect and race model inequality violation at cued locations relative to uncued locations. These results provide the first evidence of the IOR with audiovisual targets. Furthermore, IOR with audiovisual targets decreases when paying attention to both modalities. The interaction between exogenous spatial attention and audiovisual integration is discussed.

**Keywords** Inhibition of return · Audiovisual target · Audiovisual integration · Race model inequality violation · Exogenous spatial attention

---

Xiaoyu Tang and Yulin Gao contributed equally to this work and should be considered co-first authors.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00221-019-05488-0>) contains supplementary material, which is available to authorized users.

✉ Xiaoyu Tang  
tangyu-2006@163.com

✉ Ming Zhang  
psyzm@suda.edu.cn

✉ Qiong Wu  
wuqiong@okayama-u.ac.jp

<sup>1</sup> School of Psychology, Liaoning Collaborative Innovation Center of Children and Adolescents Healthy Personality Assessment and Cultivation, Liaoning Normal University, Dalian 116029, China

<sup>2</sup> Cognitive Neuroscience Laboratory, Okayama University, Okayama 7008530, Japan

<sup>3</sup> Department of Psychology, Jilin University, Changchun 130012, China

## Introduction

At every moment, we are surrounded by an overwhelming number of objects from different sensory modalities. Spatial attention plays a key role in selecting objects for

<sup>4</sup> Department of Psychology, Hubei University, Wuhan 430062, China

<sup>5</sup> Department of Psychology, Guiyang University of Chinese Medicine, Guiyang 550025, China

<sup>6</sup> Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China

<sup>7</sup> Key Laboratory of Biomimetic Robots and Systems, State Key Laboratory of Intelligent Control and Decision of Complex Systems, Beijing Institute of Technology, Beijing 100081, China

<sup>8</sup> Department of Psychology, Soochow University, Suzhou 215123, China

action or further analysis (Posner 1980). Using variants of the Posner cueing paradigm, the literature concerning the exogenous orienting of spatial attention has shown that reaction time (RT) decreases when the target appears at the same location as an abrupt peripheral onset (i.e., cued condition) compared to when the target appears at the opposite location (i.e., uncued condition). This “facilitation cuing effect”, a decreased RT in the cued condition relative to the uncued condition, is found when the cue–target stimulus onset asynchrony (SOA) is shorter than 200 ms. However, when the SOA is longer than 250 ms, an “inhibition effect”, an increased RT in the cued condition relative to the uncued condition, is found, which is termed inhibition of return (IOR) (Posner and Cohen 1984; Posner et al. 1985; Zhang et al. 2013). IOR exists in both spatial and non-spatial domains (Law et al. 1995; Zhou and Chen 2008; Chen et al. 2010). As an important mechanism of exogenous spatial orienting attention, IOR has been suggested to act in biasing attention toward novel items or spatial locations and facilitating the efficiency of visual search (Klein 1988, 2000).

The previous studies have demonstrated the spatial IOR effect in contexts of detection (Posner and Cohen 1984), localization (Pratt et al. 1997), and discrimination (Lupiáñez et al. 1997) tasks within the visual modality. IOR has also been discovered within the auditory modality (Schmidt 1996; McDonald and Ward 1999; Mayer et al. 2007), although the sensory representations of attributes of the visual and auditory modality are different (Talsma and Kok 2002). The auditory IOR can be elicited based on both location and frequency (Mondor et al. 1998; Mondor and Breau 1999). Chen and her colleagues observed the interaction between location- and frequency-based auditory IOR. They found that location-based IOR was significant only when the auditory cue and target were in the same frequency, and frequency-based IOR was significant only when the auditory cue and target were presented at the same location (Chen et al. 2007).

Investigations into spatial IOR have extended to the cross-modal link between visual and auditory information processing (see Table 1). In the cross-modal cue–target paradigm, responses to visual targets following auditory cues (Reuter-Lorenz and Rosenquist 1996) or responses to auditory targets following visual cues (Reuter-Lorenz et al. 1996) have been slowed at cued locations relative to uncued locations. A significant IOR was also found in the cross-modal target–target paradigm, in which the targets are presented successively (Spence and Driver 1998a; Spence et al. 2000). Unlike IOR within the visual or auditory modality, however, the occurrence of a cross-modal IOR is conditional (see Table 1). First, a central reorienting cue is necessary for slowing the responses to visual or auditory targets at cued locations in cross-modal contexts. The central reorienting cue can force the subject’s attention from the peripheral location back to

the central fixation. It is presented between the cue and the target in the cross-modal cue–target paradigm (Reuter-Lorenz et al. 1996; Reuter-Lorenz and Rosenquist 1996; Spence and Driver 1998b) or between the two successive targets in the cross-modal target–target paradigm (Spence and Driver 1998a; Spence et al. 2000; Roggeveen et al. 2005). Second, a longer SOA is needed to trigger the cross-modal IOR relative to the unimodal IOR. Based on the results of the previous studies (see Table 1), the occurrence of a cross-modal IOR may be delayed relative to the development of IOR within the visual or auditory modality. The shortest SOA of a significant IOR during cross-modal spatial orienting is 500 ms in the cue–target paradigm (Spence and Driver 1998b) or longer than 770 ms in the target–target paradigm (Roggeveen et al. 2005).

The occurrence of IOR within the visual/auditory domain and cross-modally is typically attributed to the neural basis of IOR, i.e., the superior colliculus (SC) (Posner et al. 1985; Anderson and Rees 2011). The SC is part of the midbrain and plays a key role in the network for spatial orienting (Kustov and Robinson 1996), receiving rich input from the visual, auditory, and tactile sensory systems (Stein and Meredith 1993). The multisensory neurons in the intermediate layers of the SC are also important for the interaction or integration of stimuli from different modalities (Meredith and Stein 1996; Wallace et al. 1998). Simultaneous inputs from both the visual and auditory modalities can influence perception and behavior (Meredith and Stein 1986; Stein et al. 1996). For example, the ventriloquism effect describes the situation in which auditory location perception (the puppet’s master) is shifted toward a visual signal (the puppet’s mouth movement) (Hairston et al. 2003). The double flash illusion shows that visual stimuli can be qualitatively altered by auditory signals (Shams et al. 2000). The redundant signals effect shows faster and more accurate responses to the simultaneous presentation of visual and auditory stimuli than responses to the unimodal stimuli (Hershenson 1962).

Although the IOR effect has been found for single visual or auditory target conditions, few studies have examined what occurs when targets from both the visual and auditory modalities are presented simultaneously. As shown in Table 1, there is only one study by Van der Stoep et al. (2017) who did not find the significant IOR for audiovisual targets, in which the participants were instructed to press a response key whenever a visual (V), auditory (A), or audiovisual (AV) target was presented. It has been suggested that simultaneous inputs from different modalities support each other in a super-additive fashion. An event-related potential (ERP) study (Talsma et al. 2007) demonstrated that the early neural activities of AV stimuli were not equal to the sum of A and (A + V) only when both the visual and auditory senses were attended. The more “salient” bimodal stimuli, compared to unimodal stimuli, can capture attention more

**Table 1** Selection of important results from the spatial inhibition of return (IOR) across different modalities

Study	Paradigm	Task	Peripheral cue modality (SOA/ms)	Central cue	Target modality	IOR effect (ms)
Van der Stoep et al. (2017)	Cue–target	Go–no-go task	V (350–450)	No	V	18*
					A	–23
Yang and Mayer (2014)	Cue–target	Location task	A (150/650) V (150/650)	No	AV	11
					V	–14***/–19**
Guerreiro et al. (2012)	Cue–target	Location task	A (100/500/1000/1500) V (100/500/1000/1500)	No	A	–57***/9
					V	–14***/–3/0/4
Roggeveen et al. (2005)	Target–target	Go–no-go task Discrimination	A (770~) V (770~)	AV	V	+***
					A	+***
Bell et al. (2004)	Cue–target	Saccade task	A (60/160/610)	No	V	–/+ ns/+ ns
Spence et al. (2000)	Target–target	Detection task	A (950–1250/1950–2250) V (950–1250/1950–2250) T (950–1250/1950–2250)	A	V	18**/11*
					T	10*/2
					A	20***/1
					T	10**/0**
Schmitt et al. (2000)	Cue–target	Detection task	A (125/175/225/575) V (125/175/225/575)	No	V	–ns/–ns–ns/–ns/
					A	+ns/+ns/+ns/ns
		Location task	A (125/175/225/575) V (125/175/225/575)		V	–ns/–ns–ns/–ns/
					A	–ns/–ns–ns/–ns/
Spence and Driver (1998b)	Cue–target	Detection task	A (500/700/900)	A	V	12***/12**/10*
				V	1/0/1	
				AV	A	16*/13*
Spence and Driver (1998a)	Target–target	Detection task	A (900, 1300/3900, 4300) V (900, 1300/3900, 4300)	No	V	17**/–2
					A	8/–2
					V	8/–5
Reuter-Lorenz et al. (1996)	Cue–target	Detection task Saccade task	V (1000, 1300)	V	A	17**/–2
					V	34**/6
Reuter-Lorenz and Rosenquist (1996)	Cue–target	Detection task	A (1000, 1300) A (1200, 1500) saccade–cue	V	V	11*
					V	13**
						–ns
						+*

A auditory, V visual, T tactile, AV audiovisual, SOA stimulus onset asynchrony represents the time interval from the cue onset to the target onset “IOR Effect” was obtained by subtracting the reaction time in the uncued location from that in the cued location, i.e., cued minus uncued. “ns” means no significance

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

intensively (Santangelo et al. 2006, 2008; Krause et al. 2012). Due to the super-additive audiovisual interaction, it can be hypothesized that the IOR effect with a bimodal audiovisual target relative to the IOR effect with a unimodal target may be decreased. This effect was demonstrated by Van der Stoep et al. (2017), in which they found no significant IOR of AV targets but significant IOR of V targets in the condition of paying attention to both modalities. However, modality-specific selective attention (attending to a modality) and divided-modality attention (attending to multiple modalities) differentially modulated audiovisual

processing (see a review of Tang et al. 2016, part 2.2). The effect of audiovisual integration on behavioral performance can be attenuated or even eliminated under conditions of modality-specific selective attention (Mozolic et al. 2008; Wu et al. 2012).

The aim of the present study was to investigate the difference between the IOR effect of AV targets and the IOR effect of V targets under conditions of modality-specific selective attention (Experiment 1) or divided-modalities attention (Experiment 2). We adapted the exogenous spatial cueing procedure for audiovisual contexts. In this procedure,

an uninformative white square is presented on the left or right side of the fixation to serve as a visual peripheral cue triggering exogenous spatial orienting. The fixation cross is extended to serve as a central cue to summon attention back to the central location. The central cue has been suggested to hasten the appearance of an IOR even at a short SOA (Pratt and Fischer 2002). We manipulated the modalities of the target stimuli (including visual, auditory, and audiovisual modalities) and compared the median RT of the targets in the cued condition to the uncued condition. Due to the decreased audiovisual integration in visual modality selective attention relative to divided-modalities attention, we hypothesized a smaller difference between the IOR with AV targets and the IOR with V targets in Experiment 1 than that in Experiment 2.

## Experiment 1

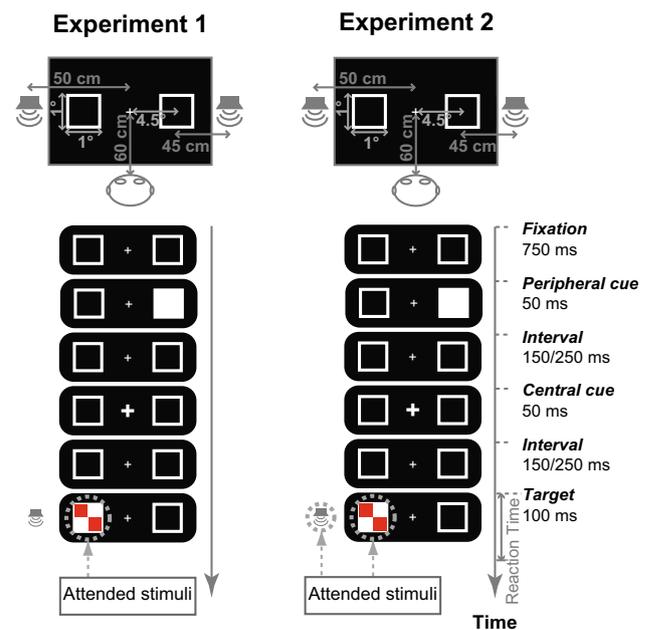
### Method

#### Participants

Twelve undergraduate students (3 females; age range: 21–27 years; mean age:  $23.3 \pm 2.2$  years) were recruited as paid volunteers. All participants were right-handed, had normal or corrected-to-normal vision, and had no history of neurological or psychiatric disorders. The volunteers had no hearing problems and had not participated in similar experiments during the previous year. The Ethics Committee of Okayama University approved the experimental protocol.

#### Stimuli

The experiment was conducted in a dimly lit, sound-attenuated room. Visual stimuli were presented on a black ( $0.4 \text{ cd/m}^2$ ) background display (with a refresh rate of 100 Hz) 60 cm from the subject (see Fig. 1). Auditory stimuli were presented via two speakers placed at both the left and the right of the display. The distance between the center of the speaker and the center of the display was 50 cm (see Fig. 1). Based on the oral reports of all of the participants, this distance can be easily judged as the left or right location. The fixation stimulus consisted of a white ( $155.2 \text{ cd/m}^2$ ) fixation cross ( $0.05^\circ \times 0.05^\circ$  of the visual angle) flanked by two white square outline boxes ( $1^\circ \times 1^\circ$ ,  $4.5^\circ$  eccentricity). Based on the fixation stimulus, one box was filled in white ( $155.2 \text{ cd/m}^2$ ) to serve as a peripheral cue and to summon attention to its location, whereas the fixation cross was extended to  $0.1^\circ \times 0.1^\circ$  of visual angle to serve as a central cue to summon attention back to the central location. There were three types of target stimuli: visual (V), auditory (A), and audiovisual (AV) targets. The V target (duration of 100 ms) was



**Fig. 1** Example of the stimuli and experimental procedure. The size and position of the stimuli are shown in the upper panel. The sequence of events and their duration (starting from the top) in the uncued audiovisual (AV) target condition are illustrated in the bottom panel. The attended (i.e., requires responses) stimuli are indicated by the dashed circle

a red ( $27.5 \text{ cd/m}^2$ ) and white ( $155.2 \text{ cd/m}^2$ ) block (subtending a visual angle of  $0.8^\circ \times 0.8^\circ$ ). The A target (duration of 100 ms) was a 1600 Hz sinusoidal tone, with linear rise and fall times of 10 ms and an intensity of 65 dB that was presented from the central speaker. The AV target consisted of the simultaneous<sup>1</sup> presentation of both the visual and the auditory stimuli.

<sup>1</sup> The E-Prime software was used in our experiments. Due to a technical problem, the presentation of the audiovisual stimulus composed of a “simultaneously” presented visual and auditory stimulus that was actually not “simultaneous”. We tested the external validation of timing of the audiovisual stimulus using a digital oscilloscope (YOKOGAWA, DL1640, 200 MHz). The E-Prime program was run on a Dell computer [OptiPlex GX520; display adapter: Intel(R) 82945G Express Chipset Family; sound card: SoundMAX Integrated Digital Audio] with a display (LG, Flatron L1751SQ-WN) and two sound speakers (SANWA SUPPLY INC, MN-SPL2). The test showed that the auditory component appeared about 10 ms (mean = 10 ms, SD = 3 ms,  $N = 20$ ) earlier than the visual component, although they were presented on the same slide and supposed to be simultaneous. A review by Vroomen and Keetels (2010) summarized that the just noticeable differences (JNDs) for intersensory temporal order can be as low as 20 ms. In our study, the temporal interval gap between a simple visual stimulus (checkboard) and auditory stimulus (sinusoidal tone) was shorter than the lowest JNDs, which can be interpreted as the lower limit of the temporal window for multisensory integration.

## Procedure and task

At the beginning of each trial, the fixation stimulus was presented for 750 ms in the center of the monitor. Following the fixation stimulus, a visual peripheral cue stimulus appeared for 50 ms at the left or right location. Then, the fixation stimulus was randomly presented for 150/250 ms, which was followed by a central cue with a duration of 50 ms. Before the target occurrence, the fixation stimulus randomly appeared again for 150/250 ms. Thus, the SOA between the peripheral cue and the target was completed in 400/600 ms. The two SOA conditions were set randomly to resist the temporal expectation of the appearance of the targets because of one fixed SOA. The target (A, V, or AV) randomly appeared (86%) for 100 ms in the left/right box or speaker or no stimulus appeared (14%) at the variable SOA. The participants were instructed to respond to the visual target stimulus at any possible location by pressing a response button (“B” on the keyboard) as quickly and accurately as possible with the index finger of their dominant hand. The participants were asked to ignore or pay no attention to the auditory stimuli from two speakers. Finally, the fixation stimulus appeared for 900 ms to allow the participants to make corresponding responses (see Fig. 1). The inter-trial interval (ITI) was randomly set to 1000–1200 ms.

## Design

Two factors of cue validity and target type were manipulated in the experiment. The cue validity condition had two levels: cued and uncued. In cued trials, the target and cue appeared at the same location. In uncued trials, they appeared at opposite locations. The amount of cued and uncued trials are the same. There were four target types: V, AV, A, and catch trials. The numbers of V, AV, and A trials were the same. Because the number of auditory alone and audiovisual trials was the same, sounds cannot predict the occurrence of a visual target. Following one practice block, each participant completed 4 experimental blocks for a total of 672 trials. The catch trials (no target appearance) compromised approximately 14% of the total trials and were embedded in the experimental procedure to discourage anticipatory guesses. Thus, there were 96 trials for each experimental condition. The participants were asked to maintain fixation on a cross as the fixation stimulus and were allowed to take a 5-min break between blocks. The total time for the experiment was approximately 40 min.

## Data recording and analysis

E-prime software (1.0 version, Neurobehavioral Systems, Inc.) was used to present the stimuli and to record the responses. For each participant, we calculated the median RT for each condition by taking the correct responses

between 150 ms and 900 ms from post-stimulus.<sup>2</sup> Based on the results that the SOA had no effect on the interaction between target modality and cue validity (full details, see Supplementary Material, concluding accuracy and RT data), the data were collapsed over the two SOA conditions. The accuracy and RT data were then compared using a two (target type: V and AV)  $\times$  2 (cue validity: cued and uncued) repeated-measures ANOVA. The Greenhouse–Geisser epsilon correction was used to correct for non-sphericity. The Bonferroni correction was applied to the post-hoc comparisons. The effect size of the partial eta-squared ( $\eta_p^2$ ) was calculated for the ANOVA. The paired *t* test (two-tailed) was used to compare the spatial cueing effect (cued vs. uncued) of V targets to the cueing effect of AV targets.

The absolute multisensory response enhancement (aMRE) and the relative multisensory response enhancement (rMRE) were calculated for each subject in each cue validity condition (cued and uncued) to investigate the amount of speedup in the bimodal condition compared with the unimodal condition (Stevenson et al. 2014; Van der Stoep et al. 2017). Van der Stoep et al. (2017) calculated the aMRE and rMRE using the following formulas:

$$\text{aMRE} = \min(\text{median}(\text{RT}_A), \text{median}(\text{RT}_V)) - \text{median}(\text{RT}_{AV}),$$

$$\text{rMRE} = \frac{\min(\text{median}(\text{RT}_A), \text{median}(\text{RT}_V)) - \text{median}(\text{RT}_{AV})}{\min(\text{median}(\text{RT}_A), \text{median}(\text{RT}_V))} \times 100\%.$$

In Experiment 1, we asked subjects to respond to V or AV and not to A. Thus, we calculated the aMRE and rMRE using only a single visual condition vs. two unimodal conditions (Experiment 2). The paired *t* test was used to compare differences in aMRE or rMRE between the different cue validity conditions. All statistical levels (i.e.,  $\alpha$  level) were set to 0.05.

## Results

### Accuracy

The accuracy (ACC, see Table 2) was entered into a two (target type: V and AV)  $\times$  2 (cue validity: cued and uncued) repeated-measures ANOVA, showing that only the main effect of the target type was significant [ $F(1, 20) = 10.04$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.33$ ]. This finding showed that the responses to the AV target (99.5%) were more accurate than those to the V target (98.3%). No other significant main effect or interactions were found. All participants made no response on catch trials as instructed.

<sup>2</sup> The RT filter interval was set to get rid of anticipatory responses and too slow responses. Simple reaction time in visual modality is at least 150 ms (Rogers 1968). The 900 ms time limit was set to make all the participants to respond as quickly and accurately as possible.

**Table 2** Average of median reaction times (RTs, in milliseconds), accuracy (ACC, percent correct), and standard deviation (SD) for all combinations of target type (A, V, and AV) and cue validity (cued and uncued) in each experiment

Target type	Validity	Experiment 1		Experiment 2	
		RT (ms)	ACC (%)	RT (ms)	ACC (%)
AV	Cued	341 ± 27	98 ± 2	309 ± 32	99 ± 2
	Uncued	305 ± 28	99 ± 2	300 ± 31	99 ± 1
V	Cued	361 ± 36	98 ± 2	395 ± 31	96 ± 6
	Uncued	326 ± 33	98 ± 2	369 ± 35	96 ± 5
A	Cued	–	–	318 ± 34	99 ± 2
	Uncued	–	–	317 ± 33	98 ± 2

## Reaction time

The average of median RT (RT rejection rate: 2%) in each condition is shown in Table 2 and Fig. 2a. The two (target modality: V and AV) × 2 (cue validity: cued and uncued) repeated-measures ANOVA showed that the main effect of the target modality was significant [ $F(1, 11) = 53.40$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.83$ ]. The responses to the audiovisual stimuli (323 ms) were faster than those to the visual stimuli (343 ms). The main effect of cue validity was also significant [ $F(1, 11) = 200.68$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.95$ ]. The results showed that the responses in the cued condition (351 ms) were slower than those in the uncued condition (315 ms), which suggested that IOR occurred. However, the interaction between the target modality and cue validity was not significant [ $F(1, 11) = 0.18$ ,  $p = 0.68$ ,  $\eta_p^2 = 0.02$ ]. The IOR effect of the V target (35 ms) was analogous to that of the AV target (36 ms).

## Multisensory response enhancement (MRE)

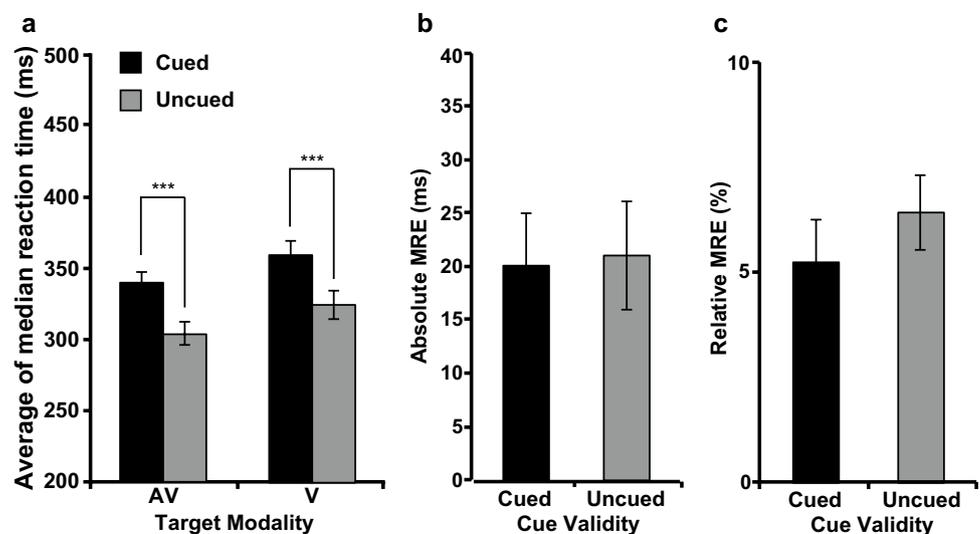
The average of aMRE and rMRE in each condition is shown in Table 3 and Fig. 2b, c. The paired  $t$  test found no significant difference on the amount of aMRE [21 ms vs. 20 ms;  $t(11) = 0.4$ ,  $p = 0.68$ ] or rMRE [6% vs. 5%;  $t(11) = 0.4$ ,  $p = 0.68$ ] between the uncued condition and the cued condition.

## Discussion

The results of Experiment 1 showed not only an IOR effect with a visual target, but also a significant IOR effect with an audiovisual target (see Table 3; Fig. 2). No significant difference between IOR with a visual target and IOR with an audiovisual target was found. The absolute and relative multisensory response enhancements were not significantly different between the cued and the uncued conditions.

We provide the first evidence (at least to our knowledge) for the IOR effect with bimodal target stimuli. This result suggests that the stimulus at the cued location can be inhibited even when the visual target is presented simultaneously with an unattended auditory signal. The present result of IOR effect with audiovisual targets is different from Van der Stoep et al. (2017) in which they did not find a significant IOR of audiovisual targets. This finding may be caused by three main variables: the central reorienting cue, the task at hand, and selectively paying attention to the visual modality. First, as we mentioned in Introduction, the central cue we used in this study forced participants' attention back to the fixation and hastened the appearance of IOR (Pratt and Fischer 2002). We thus observed a significant IOR effect even when the visual

**Fig. 2** Reaction time results in Experiment 1. **a** Average of median RT in each condition; **b** magnitude of absolute multisensory response enhancement (aMRE) in both cue validity conditions; **c** magnitude of relative multisensory response enhancement (rMRE) in both cue validity conditions. The error bars represent the standard errors of the mean (\*\*\*)  $p < 0.001$



**Table 3** Comparison of the cueing effects (ms), absolute multisensory responses enhancement (aMRE, ms), relative multisensory responses enhancement (rMRE, %), and contrasts between different conditions in each experiment employing a *t* test (sig. two-tailed, 95% confidence interval, CI)

	Mean	95% CI		<i>t</i>	Sig.
		Lower	Upper		
<i>Experiment 1</i>					
Cueing effect					
V	35	27	43	9.58	0.000
AV	36	31	42	14.16	0.000
Contrasts					
AV vs. V	2	-7	10	0.43	0.676
aMRE					
Uncued	21	15	28	7.01	0.000
Cued	20	12	28	5.32	0.000
Contrasts					
Uncued vs. cued	2	-7	10	0.43	0.676
rMRE					
Uncued	6	5	8	7.70	0.000
Cued	5	3	7	5.40	0.000
Contrasts					
Uncued vs. cued	1	-4	1	0.43	0.676
<i>Experiment 2</i>					
Cueing effect					
V	26	19	33	7.97	0.000
AV	9	5	13	5.40	0.000
A	0	-3	4	0.14	0.890
Contrasts					
AV vs. V	-17	-23	-11	-6.32	0.000
V vs. A	26	17	35	6.35	0.000
AV vs. A	9	4	13	4.43	0.001
aMRE					
Uncued	17	12	22	7.24	0.000
Cued	8	4	13	4.05	0.002
Contrasts					
Uncued vs. cued	9	4	13	4.43	0.001
rMRE					
Uncued	5	4	7	7.52	0.000
Cued	3	1	4	4.04	0.002
Contrasts					
Uncued vs. cued	3	1	4	4.48	0.001

target was accompanied by the unattended auditory signal. Second, we used a simple detection task, whereas Van der Stoep and his colleagues used a go/no-go task based on target location (i.e., only respond to peripheral target not to center target). As we know, the task at hand has effect on IOR produce (Lupiáñez et al. 1997, also see Martín-Arévalo et al. 2016), which may explain the discrepancies between our results and the results of Van der Stoep et al. (2017).

Third, in Experiment 1, the auditory stimuli were set to be unattended or ignored; attention was focused on the visual modality selectively and not divided to both visual and auditory modalities such as Van der Stoep et al., (2017). It has been suggested that paying attention to both auditory and visual modalities is a prerequisite for early audiovisual integration (Talsma et al. 2007). Furthermore, the degraded or delayed processing at the early perceptual/attentional stages has been found to be correlated with the occurrence of the IOR effect (Posner et al. 1985; Klein 1988; but see; Martín-Arévalo et al. 2016). Specifically, a biased orienting response away from the previously explored location is activated by the abrupt onset of the cue with a long SOA. The biased attention process results in fewer attentional resources to the cued location and then impairs the perceptual processing of targets at the cued location, consequently slowing manual response to the target presented at the cued location (McDonald et al. 1999, 2009; Prime and Ward 2004, 2006; Satel et al. 2013). The setting of the ignored auditory stimuli in Experiment 1 may reduce the integration of visual and auditory stimuli, which may result in the comparable magnitudes of the IOR effect for a unimodal visual target and bimodal audiovisual target was found.

Although we found comparable magnitudes of the IOR effect for a unimodal visual target and bimodal audiovisual target, the conclusion that the IOR with the audiovisual target is the same, as the IOR with the visual target is too premature. Experiment .2 was designed to confirm whether paying attention to multiple modalities had an effect on the IOR effect with an audiovisual target.

## Experiment 2

### Method

#### Participants

In this experiment, 13 undergraduate students (2 females; age range: 21–27 years; mean age: 23.1 ± 1.9 years) were recruited as paid volunteers. All participants were right-handed, had normal or corrected-to-normal vision, and had no history of neurological or psychiatric disorders. The volunteers had no hearing problems and had not participated in similar experiments during the previous year. The experimental protocol was approved by the Ethics Committee of Okayama University.

#### Task and design

The apparatus and stimuli employed were identical to those in Experiment 1. However, the participants were instructed

to respond to any detectable signals at any possible location by pressing a response button (“B” on the keyboard) as quickly and accurately as possible with the index finger of their dominant hand. Thus, participants should respond to A target, V target, and AV target. In addition, the procedure was the same as that in Experiment 1.

The experimental conditions were identical to those in Experiment 1, except that the auditory stimuli required responses as well. Following 1 practice block, each participant completed 12 experimental blocks (240 trials in each block). In addition to the three types of targets, the catch trials (no target appearance), comprising 20% of the total trials, were embedded in the experimental procedure to discourage anticipatory guesses. The participants were asked to maintain fixation on a cross in the fixation stimulus and were allowed to take a 5-min break between blocks. The total time of the experiment was approximately 90 min.

### Data analysis

The ACC and RT data were compared using a three (target type: A, V, and AV)  $\times$  2 (cue validity: cued and uncued) repeated-measures ANOVA. Using the formulas mentioned in Experiment 1, the amount of aMRE and rMRE was calculated based on the unimodal A and V conditions. The paired  $t$  test was used to compare differences in aMRE or rMRE between the different cue validity conditions.

### Race model inequality violation (differences in milliseconds for each quantile)

We used the individual cumulative distribution functions (CDFs) of each target modality in each cue validity condition to calculate the race model to demonstrate the audiovisual response enhancement effect (Miller 1982, 1986; Laurienti et al. 2006). For each participant, the predicted cumulative probability in each condition was calculated by the following formulas:

$$P(\text{RT}_{\text{Race model}} < t) = P(\text{RT}_A < t) + P(\text{RT}_V < t).$$

The observed RTs of the audiovisual CDF of each participant in each condition were compared to its corresponding race model (i.e., cued AV CDF vs. cued race model CDF; uncued AV CDF vs. uncued race model CDF) from the 10th to the 90th percentiles of each CDF to test for race model inequality (RMI) violations (Ulrich et al. 2007; Gondan and Minakata 2016). Paired  $t$  tests were used to analyze RMI violations (differences in milliseconds for each quantile,  $p$  values were Bonferroni corrected, see Van der Stoep et al. 2017). Significant violations of RMI (i.e.,  $\text{RT}_{\text{AV}} < \text{RT}_{\text{Race model}}$ ) indicate audiovisual interactions that exceed statistical facilitation (Miller 2016). To test for

differences in the amount of RMI violation between cue validity conditions, the median amount of RMI violation across the nine percentile points of the CDF of each cue validity condition of each participant was included in the paired  $t$  test (cued vs. uncued).

The positive area under the difference curve (i.e., the difference in milliseconds of  $\text{CP}_{\text{AV}}$  and  $\text{CP}_{\text{Race model}}$  at each percentile) was also used to test differences in RMI violation between cue validity conditions. To extract the positive area under the difference curve, all negative probabilities (no RMI violation) were set to a value of zero, and only the positive area under the curve was calculated for all participants (Van der Stoep et al. 2015, 2017). A paired  $t$  test (two-tailed) was used to compare the positive area under the curve between the cued condition and the uncued condition.

## Results

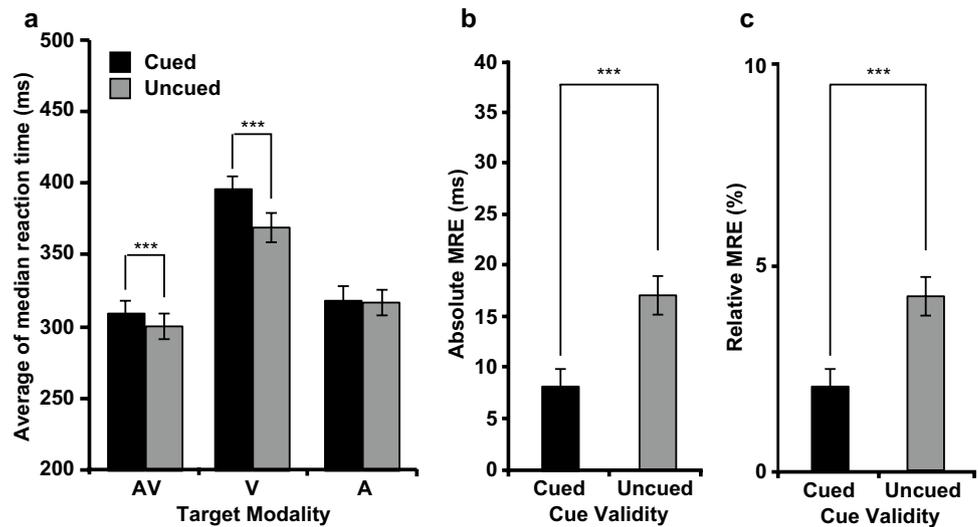
### Accuracy

The ACC data were entered into a three (target type: A, V, and AV)  $\times$  2 (cue validity: cued and uncued) repeated-measures ANOVA, showing that only the main effect of the target type was significant [ $F(2, 24) = 5.19, p = 0.04, \eta_p^2 = 0.3$ ]. This finding showed that the responses to the AV target (99%) were more accurate than those to the A (98%) or the V targets (96%). No other significant main effect or interactions were found. The accuracy of the catch trials across all participants was 96.7%.

### Reaction time

The median RTs (rejection rate: 3%) in each condition are shown in Table 2 and Fig. 3a. The RT was entered into a three (target modality: A, V, and AV)  $\times$  2 (cue validity: cued and uncued) repeated-measures ANOVA. The main effect of the target modality was significant [ $F(2, 24) = 286.29, p < 0.001, \eta_p^2 = 0.96$ ], which is driven by AV targets (305 ms) being faster than A (318 ms) and the V (382 ms) targets. The main effect of cue validity was significant [ $F(1, 12) = 60.82, p < 0.001, \eta_p^2 = 0.84$ ]. The results showed that the responses in the cued condition (341 ms) were slower than those in the uncued condition (329 ms), which suggested that an IOR occurred. In addition, the interaction between the target modality and cue validity was significant [ $F(2, 24) = 37.30, p < 0.001, \eta_p^2 = 0.76$ ]. The IOR effect was significant for both the V (26 ms,  $p < 0.001$ ) and the AV (9 ms,  $p < 0.001$ ) targets, but not for the A targets (0 ms,  $p = 0.89$ ). Further analysis found that the IOR effect of the visual stimuli was larger than that of the audiovisual stimuli [ $t(12) = 6.3, p < 0.001$ ].

**Fig. 3** Reaction time results in Experiment 2. **a** Average of median RT in each condition; **b** magnitude of absolute multisensory response enhancement (aMRE) in both cue validity conditions; **c** magnitude of relative multisensory response enhancement (rMRE) in both cue validity conditions. The error bars represent the standard errors of the mean (\*\**p* < 0.001)



The planned comparison results via paired *t* test are shown in Table 3. In addition, the IOR effect of the AV targets was even smaller than the average of the IOR effect of the V and A targets [9 ms vs. 13 ms; *t* (12) = -3.45, *p* = 0.005].

**Multisensory response enhancement (MRE)**

The average aMRE and rMRE in each condition are shown in Table 3 and Fig. 3b, c. One sample *t* tests showed significant aMRE in both cued [8 ms, *t* (12) = 4.0, *p* = 0.002] and uncued [17 ms, *t* (12) = 7.2, *p* < 0.001] conditions. The paired *t* test found larger aMRE in the uncued condition than that in the cued condition [*t* (12) = 4.4, *p* = 0.001]. Significant rMRE was observed in both cued [2.6%, *t* (12) = 4.0, *p* = 0.002] and uncued [5.3%, *t* (12) = 7.5, *p* < 0.001] conditions. The paired *t* test found a larger rMRE in the uncued condition than that in the cued condition [*t* (12) = 4.5, *p* = 0.001].

**Race model inequality violation (differences in milliseconds for each quantile)**

Figure 4a shows the amount of RMI violation in cued and uncued conditions for each of the quantiles. In the cued condition, a significant RMI violation was observed for the 30th-to-the 50th percentiles (*p* < 0.05). In the uncued condition, a significant RMI violation was observed for the 20th-to-the 60th percentiles (*p* < 0.05). The range of RTs in which the significant RMI violation was observed was broader in the uncued condition compared to the cued condition.

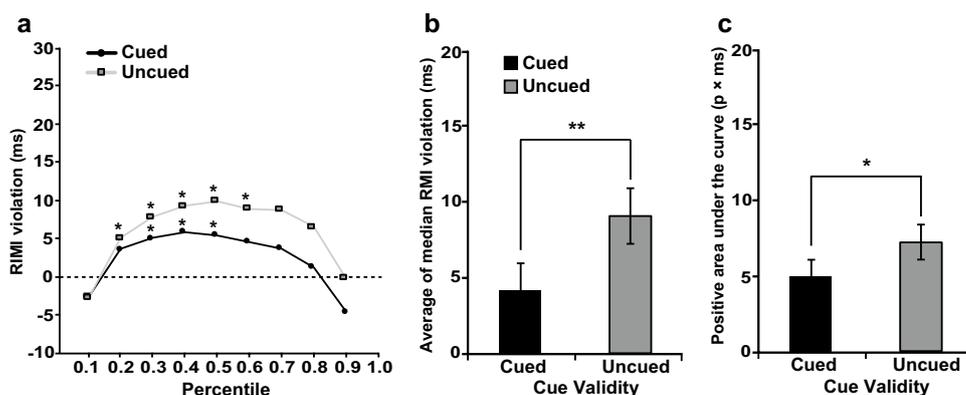
The median RMI across all percentiles was calculated and compared between cued and uncued conditions (Fig. 4b). The paired *t* test revealed that the difference in the amount of median RMI violation between the uncued and the

cued condition was significant [8 ms vs. 4 ms, *t* (12) = 3.2, *p* = 0.008]. The positive area under the curve was also compared between the two cue validity conditions (Fig. 4c). The paired *t* test showed that the difference in positive area between the uncued and cued condition was significant [7 ms vs. 5 ms, *t* (12) = 2.5, *p* = 0.019].

**Correlation between differences in unimodal RTs, rMRE, and RMI violation**

It has previously been suggested that differences in unimodal processing may have effect on multisensory integration, i.e., equal performance in different sensory modalities leading to the largest benefits of multisensory stimulation (Otto et al. 2013; Van der Stoep et al. 2017). Here, the difference between A and V RTs in the cued condition was significantly larger than that in the uncued condition [77 ms vs. 51 ms; *t* (12) = 6.3, *p* < 0.001]. To test whether the amount of rMRE and RMI violation was related to the size of the differences in unimodal RTs, we analyzed the correlation between the differences in A and V RTs and the amount of rMRE, and the correlation between the differences in A and V RTs and the average amount of RMI violation across the two cue types (cued and uncued). There was a significant correlation between the difference in A and V RTs and the amount of rMRE (Pearson *r* = -0.58, *p* = 0.037). The differences in unimodal RTs was not significantly related to the average amount of RMI violation (Pearson *r* = -0.43, *p* = 0.07). These correlations suggest that when the difference between A and V RTs was smaller, the amount of rMRE was larger.

**Fig. 4** Race model inequality violations (differences in milliseconds) results in Experiment 2. **a** Average race model inequality (RMI) violation for each percentile in both cue validity conditions; **b** average of median RMI violation across the nine percentiles in both cue validity conditions; **c** average positive area under the curve in both cue validity conditions. The error bars represent the standard errors of the mean (\* $p < 0.05$ , \*\* $p < 0.01$ )



## Discussion

The aim of Experiment 2 was to confirm whether paying attention to multiple modalities has an effect on the IOR effect with an audiovisual target. First, the results showed a significant IOR effect with both the visual and the audiovisual target but not with the auditory target. More interestingly, the magnitude of the IOR on the RTs of the unimodal visual target was larger than that of the bimodal audiovisual target (see Table 3). Second, we observed larger MRE effect and positive area under the race model inequality violation curve at uncued locations relative to cued locations (see Table 3; Figs. 3b, c and 4a, c).

### IOR with auditory target following visual cue

The result of no significant IOR effect of the unimodal auditory target following the visual cue in Experiment 2 is consistent with the previous studies (Schmitt et al. 2000; Guerreiro et al. 2012; Yang and Mayer 2014), although the evidence for an IOR in the cross-modal condition has previously been observed (Spence et al. 2000; Roggeveen et al. 2005). This finding may be caused by two experimental manipulations, i.e., the task and the SOA (Spence and Driver 1998a). First, the simple detection task used in Experiment 2 may not be appropriate for exploring orienting with auditory targets (Schmitt et al. 2000, Experiment 1; see Table 1). However, a discrimination task may be more suitable (Roggeveen et al. 2005). Second, the time course of the IOR in the cross-modal condition is longer than the time course of IOR in the unimodal condition (Tassinari and Campara 1996). Specifically, the previous studies (Table 1) found a significant IOR during cross-modal orienting (a visual cue with an auditory target) at an SOA of 1050–1350 ms (Spence et al. 2000), but not at an SOA of 350–450 ms (Van der Stoep et al. 2017), 575 ms (Schmitt et al. 2000), or 650 ms (Yang and Mayer 2014). Thus, the SOA of 400 ms or 600 ms in Experiment 2 is not long enough to observe the significant IOR effect of an auditory target following a visual cue.

Further studies employing a discrimination task or a longer SOA, such as an auditory target following a visual cue condition, are needed to confirm IOR in cross-modal orienting.

### Modulation of paying attention to multiple modalities on visually induced IOR with audiovisual targets

In Experiment 2, the participants were instructed to divide attention between both visual and auditory modalities, whereas they were instructed to selectively attend to the visual modality in Experiment 1. The previous studies showed that the effect of audiovisual integration on behavioral performance can be attenuated or even eliminated under conditions of modality-specific selective attention (Mozolic et al. 2008; Talsma et al. 2007; also see a review of; Tang et al. 2016). Hypotheses regarding the effect of modality-specific selective attention on audiovisual integration have been proposed, such as limited processing of stimuli in the unattended sensory modality (Mozolic et al. 2008). It has also been proposed that sensory gating could be changed when attention is distributed across modalities (Talsma et al. 2007; Anderson and Rees 2011). More details on the neural mechanisms of the effect on multisensory integration by modality selectivity were discussed in our review article (Tang et al. 2016). Compared with Experiment 1, Experiment 2 found a decreased IOR with audiovisual targets relative to IOR with visual targets. This result may suggest that paying attention to multiple modalities simultaneously can not only modulate audiovisual integration (Giard and Peronet 1999; Talsma and Woldorff 2005; Talsma et al. 2007), but also have effect on the interaction between audiovisual integration and exogenous attention (see more discussion in “General discussion”).

In addition, we observed a larger audiovisual integration effect (i.e., amount of median RMI violation) for the uncued condition compared to the cued condition in Experiment 2, which was consistent with the previous studies (Van der Stoep et al. 2015, 2017). As discussed above, the visually induced

IOR occurs not only for visual targets, but also not for auditory targets at the cued location, which makes a larger difference in unisensory processing times at cued locations compared to uncued locations. Differences in unimodal signal strength cause one sense to become dominant which reduces integration at cued locations compared to uncued locations (Van der Stoep et al. 2017). In Experiment 2, we also found the significant correlations between the differences in A and V RTs and the amount of rMRE. Our results supported the explanation proposed by Van der Stoep et al. (2017) that the differences in unisensory processing may reflect differences in signal strength causing one sense to become dominant so that attenuating the integration of visual and auditory information. However, what will happen if the IOR occurs for both visual and auditory targets? This question needs to be examined in the future.

Moreover, the lack of differences between the cued and the uncued unimodal auditory targets indicates that exogenous orienting to the auditory component does not contribute to the IOR effect with an audiovisual target. In other words, the decreased IOR of audiovisual targets is not due to the additive responses (in Experiment 2, we asked the participants to respond to both the auditory and visual targets), but instead is possibly due to the integration or interaction of the concurrent attended visual stimulus and attended auditory stimulus. This possibility is further discussed below.

## General discussion

The primary goal of this study was to investigate the difference between the IOR effect of AV targets and the IOR effect of V targets under conditions of modality-specific selective attention and divided-modality attention. The results showed a comparable magnitude of the IOR effect on the RTs between a unimodal visual target and a bimodal audiovisual target when selectively paying attention to the visual modality (Experiment 1). However, the IOR effect of the unimodal visual target was larger than the IOR of the bimodal audiovisual target when dividing attention between both visual and auditory modalities (Experiment 2). The results of Experiment 2 also showed a larger RMI violation (i.e., larger audiovisual integration effect) at uncued locations compared with cued locations. Taken together, the findings show that paying attention to multiple modalities influences the IOR with audiovisual targets and the interaction between audiovisual integration and exogenous spatial attention.

### Effect of exogenous spatial attention on audiovisual integration

Van der Stoep et al. (2015) provided the first evidence that exogenous spatial attention influences audiovisual

integration using a similar paradigm like ours (but investigating the facilitation effect at short SOAs). They proposed two hypotheses to explain the decreased audiovisual integration at cued location: spatial uncertainty of target location and perceptual sensitivity. The spatial uncertainty of target location is linked to the need for spatial orienting, with a higher uncertainty of the target location leading to a higher need for spatial orienting (Van der Stoep et al. 2015). The uncertainty of the target location is very high during the exogenous spatial cueing paradigm. Consequently, the spatial orienting evoked by the cue and spatial orienting caused by the audiovisual target is redundant. That is, the audiovisual integration is not as helpful in a detection task in which attention is already at the peripheral location. Therefore, audiovisual integration is attenuated at the cued location.

Van der Stoep et al. (2015) proposed another hypothesis for the decrease in audiovisual integration at cued locations compared to uncued locations: it could be related to the changes in perceptual sensitivity caused by exogenous cues. However, our study is inconsistent with this perceptual sensitivity explanation, as was Van der Stoep et al. (2017). In the short SOAs (200–250 ms) used in Van der Stoep et al. (2015), a facilitation effect with visual or auditory targets was observed. Exogenous spatial attention is able to increase perceptual sensitivity at cued locations compared to uncued locations (Carrasco et al. 2008; Carrasco 2011). The principle of inverse effectiveness (Otto et al. 2013) states that the benefit of audiovisual integration is larger for weaker stimuli than for stronger stimuli (such as stimuli presented at cued locations in short SOA). Our study (400 ms/600 ms) and Van der Stoep et al. (2017: 350–450 ms) used longer SOAs and observed an IOR effect that was associated with decreased perceptual sensitivity (not always, see Martín-Arévalo et al. 2016). Another explanation proposed by Van der Stoep (2017) states that the more differences in unimodal signal strength the weaker audiovisual integration, which was discussed in Experiment 2.

In our study, we manipulated the factor of paying attention to the visual modality alone or to both visual and auditory modalities in different experiments. We found decreased audiovisual integration at cued locations compared with uncued locations in Experiment 2. These results suggest that paying attention to multiple modalities has effect on modulation of audiovisual integration by exogenous attention. The neural mechanisms underlying this modulation effect deserve further research. Taken together, exogenous attention can modulate audiovisual integration (Van der Stoep et al. 2015; also see, Tang et al. 2016, in which different modulations of audiovisual integration by exogenous and endogenous attention were discussed).

## Effect of audiovisual integration on IOR with audiovisual targets following visual peripheral cues

Van der Stoep et al. (2015, 2017) used a similar paradigm and found the attenuated audiovisual integration at cued locations relative to uncued locations. These two studies focused on discussing the effect of exogenous attention on audiovisual integration, whereas they did not discuss the effect of audiovisual integration on IOR. As a mechanism for biasing the visual system to acquire novel information to facilitate the efficiency of visual search (Posner et al. 1985; Klein 1988, 2000), IOR has attracted a great deal of research interest. The perceptual/attentional hypothesis has been accepted by most researchers in the field, although it has been challenged by others in regard to the mechanisms underlying IOR (Berlucchi 2006; Martín-Arévalo et al. 2016). Specifically, a biased orienting response away from the previously explored location is activated by the abrupt onset of a cue with a long SOA. The biased attentional process results in fewer attentional resources to the cued location and then impairs the perceptual processing of targets at the cued location, consequently slowing the manual response to a target presented at the cued location (McDonald et al. 1999, 2009; Prime and Ward 2004, 2006; Satel et al. 2013).

In this study, the IOR effect was demonstrated on both visual targets and audiovisual targets. The IOR effect elicited by an audiovisual target was reduced when paying attention to both visual and auditory modalities (Table 3). It has been proposed that, when subjects distribute their attention to different modalities, the integration of the visual and the auditory stimuli can guide attention to the location of the concurrently presented visual stimulus (Van der Burg et al. 2008, 2011; Talsma et al. 2010). In particular, auditory signals can enhance the early perceptual processing of a synchronized visual event, make the visual event stand out from its complex environment, and increase the chances of selection of that visual event relative to the background or others distractors (Van der Burg et al. 2008, 2011). In addition, the concurrent visual event and auditory signal can not only increase the perceived visual brightness (Stein et al. 1996), but also decrease the visual contrast thresholds (Lippert et al. 2007; Noesselt et al. 2010). That is, the attended auditory stimulus can be integrated with a simultaneous visual stimulus and thus enhanced the perceptual salience<sup>3</sup> of a simultaneous visual stimulus.

<sup>3</sup> Visual salience is defined in terms of the feature contrast in any visual map. It is the distinct, subjective, perceptual quality that makes some items stand out from their background or neighbors and immediately capture attention. This type of salience plays a key role in helping the brain reasonably perform efficient selection (Van der Burg et al. 2008, 2011).

However, as discussed above, IOR can facilitate visual exploration by reducing the salience of a previously explored spatial location (i.e., the cued location), so that other unexplored locations (i.e., the uncued location) in the visual field can out-compete the explored location for focal attention—that is, favoring novel locations (Klein 1988; Prime and Ward 2006; van Koningsbruggen et al. 2010). When IOR (decreasing salience) encounters the integration of attended visual and attended auditory stimuli (increasing perceptual salience), the reduced IOR effect elicited by an audiovisual target has been found. Therefore, the present study indicates that paying attention to multiple modalities can modulate the effect of audiovisual integration on the IOR with audiovisual targets.

Moreover, in our study, the visually induced IOR effect with audiovisual targets is not eliminated, but decreases when paying attention to both visual and auditory modalities. A first possibility is that the power of enhanced perceptual salience by audiovisual integration is not sufficiently strong to be able to out-compete the reduced perceptual processing due to IOR. Alternatively, the decreased perceptual/attentional processing is not the only one component, but at least one of the components that contributes to slowed responses to an audiovisual target at the cued location (Martín-Arévalo et al. 2016). Several other mechanisms suggest that IOR may arise from inhibition of response selection (Ivanoff and Klein 2006; Prime and Jolicoeur 2009), inhibition of the motor process (Godijn and Theeuwes 2002), or a disconnect in the link between the stimulus and the response at previously attended locations (Fuentes et al. 1999; Zhang et al. 2012, 2013). Thus, further neurophysiological studies are needed to provide more direct neural evidence concerning the IOR with an audiovisual target.

The present study, together with the previous investigations, suggests that audiovisual integration/interaction has an influence on exogenous orienting of spatial attention (Tang et al. 2016). On one hand, as discussed above, audiovisual integration can modulate the perceptual salience of targets and help the audiovisual target resist being inhibited in the exogenous spatial cueing paradigm. On the other hand, audiovisual integration can act on generating the facilitation effect of exogenous spatial orienting. Specifically, rather than a unimodal cue, a multimodal cue can even trigger a significant spatial cueing effect during a high-perceptual load condition (Santangelo and Spence 2007; Santangelo et al. 2008). Similarly, multimodal cues are found to capture attention more effectively than a unimodal cue when participants are asked to complete a visual search task (Matusz and Eimer 2011). In addition, these multimodal cues can even enlarge the attentional cueing effect in visual–spatial working memory compared to unimodal cues (Botta et al. 2011).

## Summary

To the best of our knowledge, the present study provides the first empirical evidence of IOR with audiovisual targets presented after visual peripheral cues. The IOR effect elicited by an audiovisual target is reduced when paying attention to both visual and auditory modalities, which indicates that audiovisual integration can affect IOR specifically when attention is distributed across the senses. We also found reduced audiovisual integration at cued locations compared with uncued locations when paying attention to multiple modalities, which support the idea that exogenous spatial attention can influence audiovisual integration. Taken together, bimodal-divided attention can modulate the interaction between exogenous spatial attention and audiovisual integration. The suggestion sheds new light on the interplay between audiovisual integration and attention. As proposed in our previous review study (Tang et al. 2016), audiovisual integration can interact not only with top-down goal-driven attentional control, but also with the bottom-up, stimulus-driven exogenous orienting of spatial attention (Van der Stoep et al. 2015).

**Acknowledgements** This study was supported in part by a Grant-in-Aid for Strategic Research Promotion from Okayama University, Japan Society for the Promotion of Science (25249026 and 25303013, JW), National Natural Science Foundation of China (31600882, XT; 61473043, JW; 31371025 and 31871092, MZ; 31700973, WY; 31800932, YR), Fundamental Research Funds for the Central Universities (450060502081, YG); Humanity and Social Science Youth Foundation of Ministry of Education of China (16YJC190025, WY; 18XJC190003, YR).

## Compliance with ethical standards

**Conflict of interest** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Anderson EJ, Rees G (2011) Neural correlates of spatial orienting in the human superior colliculus. *J Neurophysiol* 106:2273–2284. <https://doi.org/10.1152/jn.00286.2011>
- Berlucchi G (2006) Inhibition of return: a phenomenon in search of a mechanism and a better name. *Cogn Neuropsychol* 23:1065–1074. <https://doi.org/10.1080/02643290600588426>
- Bell AH, Fecteau JH, Munoz DP (2004) Using auditory and visual stimuli to investigate the behavioral and neuronal consequences of reflexive covert orienting. *J Neurophysiol* 91:2172–2084. <https://doi.org/10.1152/jn.01080.2003>
- Botta F, Santangelo V, Raffone A, Sanabria D, Lupianez J, Belardinelli MO (2011) Multisensory integration affects visuo-spatial working memory. *J Exp Psychol Hum* 37:1099–1109. <https://doi.org/10.1037/A0023513>
- Carrasco M (2011) Visual attention: the past 25 years. *Vis Res* 51:1484–1525. <https://doi.org/10.1016/j.visres.2011.04.012>
- Carrasco M, Fuller S, Ling S (2008) Transient attention does increase perceived contrast of suprathreshold stimuli: a reply to Prinzmetal, Long and Leonhardt (2008). *Percept Psychophys* 70:1151–1164. <https://doi.org/10.3758/PP.70.7.1151>
- Chen Q, Zhang M, Zhou X (2007) Interaction between location- and frequency-based inhibition of return in human auditory system. *Exp Brain Res* 176:630–640. <https://doi.org/10.1007/s00221-006-0642-0>
- Chen Q, Fuentes LJ, Zhou X (2010) Biasing the organism for novelty: a pervasive property of the attention system. *Hum Brain Mapp* 31:1146–1156. <https://doi.org/10.1002/hbm.20924>
- Fuentes LJ, Vivas AB, Humphreys GW (1999) Inhibitory mechanisms of attentional networks: spatial and semantic inhibitory processing. *J Exp Psychol Hum* 25:1114–1126. <https://doi.org/10.1037/0096-1523.25.4.1114>
- Giard M, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* 11:473–490. <https://doi.org/10.1162/089892999563544>
- Godijn R, Theeuwes J (2002) Oculomotor capture and inhibition of return: evidence for an oculomotor suppression account of IOR. *Psychol Res* 66:234–246. <https://doi.org/10.1007/s00426-002-0098-1>
- Gondan M, Minakata K (2016) A tutorial on testing the race model inequality. *Atten Percept Psychophys* 78:723–735. <https://doi.org/10.3758/s13414-015-1018-y>
- Guerreiro MJ, Adam JJ, Van Gerven PW (2012) Automatic selective attention as a function of sensory modality in aging. *J Gerontol B Psychol* 67:194–202. <https://doi.org/10.1093/geronb/gbr090>
- Hairston WD, Wallace MT, Vaughan JW, Stein BE, Norris JL, Schirillo JA (2003) Visual localization ability influences cross-modal bias. *J Cogn Neurosci* 15:20–29. <https://doi.org/10.1162/089892903321107792>
- Hershenson M (1962) Reaction time as a measure of intersensory facilitation. *J Exp Psychol* 63:289–293. <https://doi.org/10.1037/h0039516>
- Ivanoff J, Klein RM (2006) Inhibition of return: Sensitivity and criterion as a function of response time. *J Exp Psychol Hum* 32:908. <https://doi.org/10.1037/0096-1523.32.4.908>
- Klein RM (1988) Inhibitory tagging system facilitates visual search. *Nature* 334:430–431. <https://doi.org/10.1038/334430a0>
- Klein RM (2000) Inhibition of return. *Trends Cogn Sci* 4:138–147. [https://doi.org/10.1016/S1364-6613\(00\)01452-2](https://doi.org/10.1016/S1364-6613(00)01452-2)
- Krause H, Schneider TR, Engel AK, Senkowski D (2012) Capture of visual attention interferes with multisensory speech processing. *Front Integr Neurosci* 6:1–8. <https://doi.org/10.3389/fnint.2012.00067>
- Kustov AA, Robinson DL (1996) Shared neural control of attentional shifts and eye movements. *Nature* 384:74–77. <https://doi.org/10.1038/384074a0>
- Laurienti PJ, Burdette JH, Maldjian JA, Wallace MT (2006) Enhanced multisensory integration in older adults. *Neurobiol Aging* 27:1155–1163. <https://doi.org/10.1016/j.neurobiolaging.2005.05.024>
- Law MB, Pratt J, Abrams RA (1995) Color-based inhibition of return. *Percept Psychophys* 57:402–408. <https://doi.org/10.3758/BF03213064>
- Lippert M, Logothetis NK, Kayser C (2007) Improvement of visual contrast detection by a simultaneous sound. *Brain Res* 1173:102–109. <https://doi.org/10.1016/j.brainres.2007.07.050>
- Lupiáñez J, Milán EG, Tornay FJ, Madrid E, Tudela P (1997) Does IOR occur in discrimination tasks? Yes, it does, but later. *Atten Percept Psychophys* 59:1241–1254. <https://doi.org/10.3758/BF03214211>
- Martín-Arvalo E, Chica AB, Lupiáñez J (2016) No single electrophysiological marker for facilitation and inhibition of return: a

- review. *Behav Brain Res* 300:1–10. <https://doi.org/10.1016/j.bbr.2015.11.030>
- Matusz PJ, Eimer M (2011) Multisensory enhancement of attentional capture in visual search. *Psychon B Rev* 18:904–909. <https://doi.org/10.3758/s13423-011-0131-8>
- Mayer AR, Harrington DL, Stephen J, Adair JC, Lee RR (2007) An event-related fMRI Study of exogenous facilitation and inhibition of return in the auditory modality. *J Cogn Neurosci* 19:455–467. <https://doi.org/10.1162/jocn.2007.19.3.455>
- McDonald JJ, Ward LM (1999) Spatial relevance determines facilitatory and inhibitory effects of auditory covert spatial orienting. *J Exp Psychol Hum* 25:1234–1252. <https://doi.org/10.1037/0096-1523.25.5.1234>
- McDonald JJ, Ward LM, Kiehl KA (1999) An event-related brain potential study of inhibition of return. *Percept Psychophys* 61:1411–1423. <https://doi.org/10.3758/BF03206190>
- McDonald JJ, Hickey C, Green JJ, Whitman JC (2009) Inhibition of return in the covert deployment of attention: evidence from human electrophysiology. *J Cogn Neurosci* 21:725–733. <https://doi.org/10.1162/jocn.2009.21042>
- Meredith MA, Stein BE (1986) Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res* 365:350–354. [https://doi.org/10.1016/0006-8993\(86\)91648-3](https://doi.org/10.1016/0006-8993(86)91648-3)
- Meredith MA, Stein BE (1996) Spatial determinants of multisensory integration in cat superior colliculus neurons. *J Neurophysiol* 75:1843–1857. <https://doi.org/10.1152/jn.1996.75.5.1843>
- Miller J (1982) Divided attention: evidence for coactivation with redundant signals. *Cogn Psychol* 14:247–279. [https://doi.org/10.1016/0010-0285\(82\)90010-X](https://doi.org/10.1016/0010-0285(82)90010-X)
- Miller J (1986) Timecourse of coactivation in bimodal divided attention. *Percept Psychophys* 40:331–343. <https://doi.org/10.3758/BF03203025>
- Miller J (2016) Statistical facilitation and the redundant signals effect: what are race and coactivation models? *Atten Percept Psychophys* 78:516–519. <https://doi.org/10.3758/s13414-015-1017-z>
- Mondor TA, Breau LM (1999) Facilitative and inhibitory effects of location and frequency cues: evidence of a modulation in perceptual sensitivity. *Percept Psychophys* 61:438–444. <https://doi.org/10.3758/BF03211964>
- Mondor TA, Breau LM, Milliken B (1998) Inhibitory processes in auditory selective attention: evidence of location-based and frequency-based inhibition of return. *Percept Psychophys* 60:296–302. <https://doi.org/10.3758/BF03206038>
- Mozolic JL, Hugenschmidt CE, Peiffer AM, Laurienti PJ (2008) Modality-specific selective attention attenuates multisensory integration. *Exp Brain Res* 184:39–52. <https://doi.org/10.1007/s00221-007-1080-3>
- Noesselt T, Tyll S, Boehler CN, Budinger E, Heinze HJ, Driver J (2010) Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. *J Neurosci* 30:13609–13623. <https://doi.org/10.1523/Jneurosci.4524-09.2010>
- Otto TU, Dassy B, Mamassian P (2013) Principles of multisensory behavior. *J Neurosci* 33:7463–7474
- Posner MI (1980) Orienting of attention. *Q J Exp Psychol* 32:3–25. <https://doi.org/10.1080/0033558008248231>
- Posner MI, Cohen Y (1984) Components of visual orienting. In: Bouma H, Bouwhuis DG (eds) *Attention & performance X: control of language processes*. Psychology Press Erlbaum, Hillsdale, pp 531–556
- Posner MI, Rafal RD, Choate LS, Vaughan J (1985) Inhibition of return: neural basis and function. *Cogn Neuropsychiatry* 2:211–228
- Pratt J, Fischer MH (2002) Examining the role of the fixation cue in inhibition of return. *Can J Psychol* 56:294. <https://doi.org/10.1037/h0087405>
- Pratt J, Kingstone A, Khoe W (1997) Inhibition of return in location- and identity-based choice decision tasks. *Percept Psychophys* 59:964–971. <https://doi.org/10.3758/BF03205511>
- Prime DJ, Jolicoeur P (2009) Response-selection conflict contributes to inhibition of return. *J Cogn Neurosci* 21:991–999. <https://doi.org/10.1162/jocn.2009.21105>
- Prime DJ, Ward LM (2004) Inhibition of return from stimulus to response. *Psychol Sci* 15:272. <https://doi.org/10.1111/j.0956-7976.2004.00665.x>
- Prime DJ, Ward LM (2006) Cortical expressions of inhibition of return. *Brain Res* 1072:161–174. <https://doi.org/10.1016/j.brainres.2005.11.081>
- Reuter-Lorenz PA, Rosenquist JN (1996) Auditory cues and inhibition of return: the importance of oculomotor activation. *Exp Brain Res* 112:119–126. <https://doi.org/10.1007/BF00227185>
- Reuter-Lorenz PA, Jha AP, Rosenquist JN (1996) What is inhibited in inhibition of return. *J Exp Psychol Hum* 22:367. <https://doi.org/10.1037/0096-1523.22.2.367>
- Rogers E (1968) Simple visual and simple auditory reaction time: a comparison. *Psychon Sci* 10:335–336. <https://doi.org/10.3758/BF03331548>
- Roggeveen AB, Prime DJ, Ward LM (2005) Inhibition of return and response repetition within and between modalities. *Exp Brain Res* 167:86–94. <https://doi.org/10.1007/s00221-005-0010-5>
- Santangelo V, Spence C (2007) Multisensory cues capture spatial attention regardless of perceptual load. *J Exp Psychol Hum* 33:1311–1321. <https://doi.org/10.1037/0096-1523.33.6.1311>
- Santangelo V, Van der Lubbe RH, Belardinelli MO, Postma A (2006) Spatial attention triggered by unimodal, crossmodal, and bimodal exogenous cues: a comparison of reflexive orienting mechanisms. *Exp Brain Res* 173:40–48. <https://doi.org/10.1007/s00221-006-0361-6>
- Santangelo V, Ho C, Spence C (2008) Capturing spatial attention with multisensory cues. *Psychon B Rev* 15:398–403. <https://doi.org/10.3758/PBR.15.2.398>
- Satell J, Hilchey MD, Wang ZG, Story R, Klein RM (2013) The effects of ignored versus foveated cues upon inhibition of return: an event-related potential study. *Atten Percept Psychophys* 75:29–40. <https://doi.org/10.3758/s13414-012-0381-1>
- Schmidt WC (1996) ‘Inhibition of return’ without visual input. *Neuropsychologia* 34:943–952. [https://doi.org/10.1016/0028-3932\(96\)00015-2](https://doi.org/10.1016/0028-3932(96)00015-2)
- Schmitt M, Postma A, Haan ED (2000) Interactions between exogenous auditory and visual spatial attention. *Q J Exp Psychol A* 53:105–130. <https://doi.org/10.1080/713755882>
- Shams L, Kamitani Y, Shimojo S (2000) Illusions—what you see is what you hear. *Nature* 408:788–788. <https://doi.org/10.1038/35048669>
- Spence C, Driver J (1998a) Auditory and audiovisual inhibition of return. *Percept Psychophys* 60:125–139. <https://doi.org/10.3758/BF03211923>
- Spence C, Driver J (1998b) Inhibition of return following an auditory cue. The role of central reorienting events. *Exp Brain Res* 118:352–360. <https://doi.org/10.1007/s002210050289>
- Spence C, Lloyd D, McGlone F, Nicholls MER, Driver J (2000) Inhibition of return is supramodal: a demonstration between all possible pairings of vision, touch, and audition. *Exp Brain Res* 134:42–48. <https://doi.org/10.1007/s002210000442>
- Stein BE, Meredith MA (1993) *The merging of the senses*. The MIT Press, Cambridge
- Stein BE, London N, Wilkinson LK, Price DD (1996) Enhancement of perceived visual intensity by auditory stimuli: a psychophysical

- analysis. *J Cogn Neurosci* 8:497–506. <https://doi.org/10.1162/jocn.1996.8.6.497>
- Stevenson RA, Ghose D, Fister JK et al (2014) Identifying and quantifying multisensory integration: a tutorial review. *Brain Topogr* 27:707–730. <https://doi.org/10.1007/s10548-014-0365-7>
- Talsma D, Kok A (2002) Intermodal spatial attention differs between vision and audition: an event-related potential analysis. *Psychophysiology* 39:689–706. <https://doi.org/10.1017/S004857720200152x>
- Talsma D, Woldorff MG (2005) Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *J Cogn Neurosci* 17:1098–1114. <https://doi.org/10.1162/089829054475172>
- Talsma D, Doty TJ, Woldorff MG (2007) Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb Cortex* 17:679–690. <https://doi.org/10.1093/cercor/bhk016>
- Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between attention and multisensory integration. *Trends Cogn Sci* 14:400–410. <https://doi.org/10.1016/j.tics.2010.06.008>
- Tang X, Wu J, Shen Y (2016) The interactions of multisensory integration with endogenous and exogenous attention. *Neurosci Biobehav Rev* 61:208–224. <https://doi.org/10.1016/j.neubiorev.2015.11.002>
- Tassinari G, Campara D (1996) Consequences of covert orienting to non-informative stimuli of different modalities: a unitary mechanism? *Neuropsychologia* 34:235–245. [https://doi.org/10.1016/0028-3932\(95\)00085-2](https://doi.org/10.1016/0028-3932(95)00085-2)
- Ulrich R, Miller J, Schröter H (2007) Testing the race model inequality: an algorithm and computer programs. *Behav Res Methods* 39:291–302. <https://doi.org/10.3758/BF03193160>
- van Koningsbruggen MG, Gabay S, Sapir A, Henik A, Rafal RD (2010) Hemispheric asymmetry in the remapping and maintenance of visual saliency maps: a TMS study. *J Cogn Neurosci* 22:1730–1738. <https://doi.org/10.1162/jocn.2009.21356>
- Van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J (2008) Pip and pop: nonspatial auditory signals improve spatial visual search. *J Exp Psychol Hum* 34:1053–1065. [https://doi.org/10.1016/0028-3932\(95\)00085-2](https://doi.org/10.1016/0028-3932(95)00085-2)
- Van der Burg E, Talsma D, Olivers CNL, Hickey C, Theeuwes J (2011) Early multisensory interactions affect the competition among multiple visual objects. *Neuroimage* 55:1208–1218. <https://doi.org/10.1016/j.neuroimage.2010.12.068>
- Van der Stoep N, Van der Stigchel S, Nijboer TC (2015) Exogenous spatial attention decreases audiovisual integration. *Atten Percept Psychophys* 77:464–482. <https://doi.org/10.3758/s13414-014-0785-1>
- Van der Stoep N, Van der Stigchel S, Nijboer TC, Spence C (2017) Visually induced inhibition of return affects the integration of auditory and visual information. *Perception* 46:6–17. <https://doi.org/10.1177/0301006616661934>
- Vroomen J, Keetels M (2010) Perception of intersensory synchrony: a tutorial review. *Atten Percept Psychophys* 72:871–884. <https://doi.org/10.3758/APP.72.4.871>
- Wallace MT, Meredith MA, Stein BE (1998) Multisensory integration in the superior colliculus of the alert cat. *J Neurophysiol* 80:1006–1010. <https://doi.org/10.1152/jn.1998.80.2.1006>
- Wu J, Yang J, Yu Y et al (2012) Delayed audiovisual integration of patients with mild cognitive impairment and Alzheimer's disease compared with normal aged controls. *J Alzheimer Dis* 32:317–328. <https://doi.org/10.3233/JAD-2012-111070>
- Yang Z, Mayer AR (2014) An event-related fMRI study of exogenous orienting across vision and audition. *Hum Brain Mapp* 35:964–974. <https://doi.org/10.1002/hbm.22227>
- Zhang Y, Zhou X, Zhang M (2012) Temporary inhibitory tagging at previously attended locations: evidence from event-related potentials. *Psychophysiology* 49:1191–1199. <https://doi.org/10.1111/j.1469-8986.2012.01412.x>
- Zhang M, Tang X, Wu J (2013) Blocking the link between stimulus and response at previously attended locations: evidence for inhibitory tagging mechanism. *Neurosci Biomed Eng* 1:13–21. <https://doi.org/10.1111/j.1469-8986.2012.01412.x>
- Zhou X, Chen Q (2008) Neural correlates of spatial and non-spatial inhibition of return (IOR) in attentional orienting. *Neuropsychologia* 46:2766–2775. <https://doi.org/10.1016/j.neuropsychologia.2008.05.017>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.