The Effect of Inhibition of Return on Audio-Visual Cross-Modal Correspondence (Post-Print)

Authors: Zu Guangyao, Li Shuqi, Zhang Tianyang, Wang Aijun, Zhang Ming

Date: 2023-10-09T00:00:00+00:00

Abstract

Audiovisual crossmodal correspondence has been widely observed across different types of visual and auditory stimuli, but its stage of occurrence remains unclear. This study employed a cue-target paradigm to investigate the influence of inhibition of return (IOR) on audiovisual crossmodal correspondence. Experiment 1 manipulated the spatial consistency between cue and target and the crossmodal correspondence consistency between auditory pitch and visual target location. The results revealed an interaction between the IOR effect and audiovisual crossmodal correspondence: a stable audiovisual crossmodal correspondence effect was present at cued locations, whereas the effect disappeared at uncued locations. Experiment 2 manipulated whether an irrelevant auditory stimulus was presented and found no interaction between the IOR effect and the mere presence or absence of sound, ruling out confounding effects of alerting on the results. Experiment 3 extended the stimulus onset asynchrony (SOA) between cue and target, finding that as the IOR effect weakened, the audiovisual crossmodal correspondence effect at cued locations also weakened accordingly, and the modulatory effect of IOR on audiovisual crossmodal correspondence diminished. The study demonstrates that only when crossmodal correspondence occurs between auditory stimuli and visual spatial locations does it interact with the IOR effect, which also occurs at the perceptual level, supporting that audiovisual crossmodal correspondence occurs at the perceptual stage. Furthermore, the results support the principle of inverse effectiveness for the occurrence of audiovisual crossmodal correspondence.

Full Text

The Effect of Inhibition of Return on Audiovisual Cross-Modal Correspondence

Acta Psychologica Sinica 2023, Vol. 55, No. 8, 1220–1233

https://doi.org/10.3724/SP.J.1041.2023.01220

ZU Guangyao[1], LI Shuqi[1], ZHANG Tianyang[2], WANG Aijun[1], ZHANG Ming[1]
[1] Department of Psychology, Research Center for Psychology and Behavioral Sciences, Soochow University, Suzhou 215123, China
[2] School of Public Health, Soochow University, Suzhou 215123, China

Abstract

Audiovisual cross-modal correspondence has been widely observed across different types of visual and auditory stimuli, yet the stage at which it occurs remains unclear. This study employed a cue-target paradigm to investigate the influence of inhibition of return (IOR) on audiovisual cross-modal correspondence.

Experiment 1 manipulated both the spatial consistency between cue and target and the cross-modal correspondence consistency between auditory pitch and visual target location. The results revealed an interaction between IOR and audiovisual cross-modal correspondence: a stable cross-modal correspondence effect was observed at cued locations, while this effect disappeared at uncued locations. Experiment 2 manipulated whether an irrelevant auditory stimulus was presented, finding no interaction between IOR and the mere presence of a sound, thereby ruling out alerting effects as a confounding factor. Experiment 3 extended the stimulus onset asynchrony (SOA) between cue and target, demonstrating that as the IOR effect diminished, the cross-modal correspondence effect at cued locations also weakened, and the modulatory effect of IOR on cross-modal correspondence decreased.

These findings indicate that only when auditory stimuli and visual spatial positions engage in cross-modal correspondence does an interaction occur with the IOR effect, which also occurs at the perceptual level, supporting the hypothesis that audiovisual cross-modal correspondence arises during the perceptual stage. The results also support the principle of inverse effectiveness for the occurrence of audiovisual cross-modal correspondence.

Keywords: audiovisual cross-modal correspondence, inhibition of return, cue-target paradigm, alerting effect
Classification Code: B842

Introduction

Humans rely on multiple sensory modalities to perceive their environment, and the integration of signals across different sensory channels enhances behavioral responses, a phenomenon known as multisensory response enhancement (Frassinetti et al., 2002; Stein et al., 1989). Previous research on multisensory response enhancement has primarily focused on multisensory integration (McCracken et al., 2019; Starke et al., 2017). Multisensory integration refers to the process by which individuals combine information from different sensory channels to form coherent and meaningful representations when such information exhibits spatiotemporal proximity (Peng et al., 2019; Tang et al., 2016). The integration of visual and auditory input, known as audiovisual integration, produces a redundant effect that facilitates the detection and discrimination of bimodal stimuli compared with unimodal visual or auditory stimuli (Peng et al., 2019; Tang et al., 2020; Stein & Stanford, 2008; Talsma & Woldorff, 2005). For example, visual and auditory stimuli presented simultaneously at the same location provide redundant spatial and temporal information, enhancing individuals' response capabilities (Spence, 2013).

Previous studies have found that visual and auditory stimuli can influence participants' responses not only in a redundant manner but also in a non-redundant fashion, where the two types of stimuli provide information along different dimensions yet map onto each other to influence current behavior. This phenomenon is termed audiovisual cross-modal correspondence (Spence, 2011, 2019). A common example of audiovisual cross-modal correspondence involves the mapping between auditory pitch and visual spatial location, wherein individuals tend to associate high-pitched sounds with high spatial positions and low-pitched sounds with low spatial positions. When a high-pitched sound accompanies or precedes a visual stimulus, participants respond faster to visual stimuli presented at high spatial locations than to those at low spatial locations, and vice versa (Chiou & Rich, 2012; Evans & Treisman, 2010; McCormick et al., 2018; Spence, 2019; Zeljko et al., 2019). In addition, cross-modal correspondences have been observed between pitch and stimulus size (Brunetti et al., 2018; Parise & Spence, 2008), pitch and brightness (Maimon et al., 2020; Marks et al., 2003), and pitch and spatial frequency (Evans & Treisman, 2010). Unlike audiovisual integration, which requires visual and auditory stimuli to be presented in close spatiotemporal proximity (Spence, 2013), audiovisual cross-modal correspondence can occur when visual and auditory stimuli are presented at different locations and with relatively longer temporal intervals (Chiou & Rich, 2012). Audiovisual cross-modal correspondence represents a relative mapping (Chiou & Rich, 2012); for instance, in the cross-modal correspondence between pitch and spatial location, there is no absolute mapping between a specific frequency and a particular height in space. Instead, the mapping exists between the relatively higher or lower of two tones and the relatively higher or lower of two spatial positions.

Current research presents divergent views regarding the stage at which audiovi-

sual cross-modal correspondence occurs. The first perspective posits that audiovisual cross-modal correspondence arises at the perceptual level, enhancing the perceptual saliency of stimuli (Evans & Treisman, 2010). Studies have found that when participants view ambiguously moving gratings, they are more likely to perceive upward motion when accompanied by rising pitch and downward motion when accompanied by falling pitch (Maeda et al., 2004). ERP research has demonstrated that when a preceding sound is congruent with the current visual symbol, the amplitude of the early perceptual component N1 evoked by vision is larger than when the correspondence is incongruent, supporting the view that audiovisual cross-modal correspondence reflects perceptual enhancement (Kovi ć et al., 2010). The second perspective suggests that audiovisual cross-modal correspondence occurs at the semantic level (Spence, 2011). Researchers have noted that in nearly all languages, people describe different sound frequencies using the terms "high" and "low," which also correspond to spatial positions (high and low) and spatial frequencies (high and low). This shared semantic coding across visual and auditory dimensions is thought to underlie audiovisual cross-modal correspondence. Studies using the spoken words "high" and "low" instead of actual high and low frequency tones have found cross-modal correspondence between speech and spatial location, supporting the semantic-level account (Gallace & Spence, 2006). The third perspective argues that audiovisual cross-modal correspondence occurs at a later decision-making stage (Spence, 2011), where correspondence between modalities lowers the response criterion for presented target stimuli. Using signal detection theory, researchers have found that while audiovisual cross-modal correspondence does not affect perceptual sensitivity (d'), it increases false alarm rates for targets. The argument is that if audiovisual cross-modal correspondence occurred at the perceptual level, the enhancement of response capability would not be accompanied by increased false alarm rates, thus suggesting it occurs at the decision rather than perceptual level (Marks et al., 2003). In summary, due to differences in stimulus materials and measurement indices, the stage at which audiovisual cross-modal correspondence occurs remains unresolved (Spence, 2011).

Visual and auditory integration or correspondence can enhance target perceptual saliency (Evans & Treisman, 2010; Kovi ć et al., 2010; Tang et al., 2019) and facilitate behavioral responses, while inhibition of return (IOR) in the attentional system also influences human perception of targets (Tang et al., 2020; Tang et al., 2019). IOR refers to the phenomenon in a cue-target paradigm where, when the SOA between cue and target exceeds approximately 300 ms, participants respond more slowly to stimuli appearing at the cued location than to those at uncued locations (Posner & Cohen, 1984). IOR prevents repeated search of the same location and improves visual search efficiency (Redden et al., 2021). Although different theoretical explanations exist for the mechanism of IOR, it is widely believed that IOR reflects perceptual inhibition: attention disengages from the cued location, reducing the perceptual saliency of targets at that location and impairing responses (Klein, 2000; Satel et al., 2013). ERP studies have shown that during IOR, early visual components P1 and N1 evoked

by stimuli at cued locations have lower amplitudes than those at uncued locations (Hopfinger & Mangun, 2001; Prime & Jolicoeur, 2009), supporting the notion that the IOR effect occurs at early perceptual stages.

Previous research has examined interactions between IOR and multisensory stimuli, primarily focusing on audiovisual integration (Peng et al., 2019; Tang et al., 2019; van der Stoep, van der Stigchel, et al., 2015). Researchers using cue-target paradigms with audiovisual targets have found that IOR modulates audiovisual integration. Some studies have reported smaller audiovisual integration effects at cued locations (Peng et al., 2019; Tang et al., 2019; van der Stoep et al., 2016), while others have found the opposite result—larger audiovisual integration effects at cued locations (Tang et al., 2020). These divergent findings may be related to different SOA settings across experiments (Tang et al., 2020), but existing research consistently indicates that audiovisual integration occurs at the perceptual stage (Tang et al., 2019) and is therefore subject to modulation by the IOR effect, which also occurs at the perceptual processing stage (Peng et al., 2019; Tang et al., 2020).

Although audiovisual integration and audiovisual cross-modal correspondence enhance multisensory responses through different mechanisms—the former by providing redundant information through spatiotemporal proximity (Noesselt et al., 2007; Santangelo et al., 2008) and the latter through cross-dimensional mapping that facilitates behavior (Chiou & Rich, 2012; McCormick et al., 2018)—if audiovisual cross-modal correspondence occurs at the perceptual stage, it should be influenced by the IOR effect, which occurs at the same processing stage, resulting in an interaction between the two effects. According to the additive factors logic of reaction time experiments (Sternberg, 1969), when IOR occurs, the reduced perceptual saliency of targets at cued locations should affect audiovisual cross-modal correspondence. Conversely, if audiovisual cross-modal correspondence occurs at the semantic or decision-making level, IOR should not influence it. Therefore, the present study combined spatial cueing paradigms with audiovisual cross-modal correspondence tasks to investigate the relationship between IOR and audiovisual cross-modal correspondence.

In previous audiovisual cross-modal correspondence research, visual and auditory stimuli were presented simultaneously (Brunel et al., 2015; Gallace & Spence, 2006; Getz & Kubovy, 2018), meaning that measured outcomes likely reflected both audiovisual integration and audiovisual cross-modal correspondence. Given that spatiotemporal proximity is a necessary condition for audiovisual integration (Spence, 2011) and that research has shown audiovisual integration effects disappear when the interval between visual and auditory stimuli exceeds 100 ms (van der Stoep, Spence, et al., 2015), the present study presented visual targets 200 ms after the offset of auditory stimuli and used binaural presentation to minimize audiovisual integration effects.

Additionally, auditory stimuli presented before visual targets can produce an alerting effect (Wiegand & Sander, 2019), which enhances perceptual capacity for visual stimuli (Kusnir et al., 2011) and interacts with target perceptual

saliency (Botta et al., 2017). Therefore, alerting effects may interact with exogenous cue-induced IOR effects, creating different alerting effects across conditions. Given that this is the first study to investigate the relationship between IOR and audiovisual cross-modal correspondence, it is essential to exclude potential confounding factors. Thus, we designed a control experiment to rule out possible interactions between alerting effects and IOR that could confound the results. Finally, to further explore the mechanism underlying the interaction between IOR and audiovisual cross-modal correspondence, we manipulated the SOA between cue and target to vary the magnitude of the IOR effect (Lupiáñez et al., 1997) and examined how IOR magnitude influences audiovisual cross-modal correspondence.

According to the principle of inverse effectiveness in multisensory response enhancement (Meredith & Stein, 1983; van der Stoep et al., 2016), we hypothesized that as SOA increases, the IOR effect would decrease (Lupiáñez et al., 1997), leading to a reduction in the cross-modal correspondence effect at cued locations and a weakening of IOR's modulatory effect on audiovisual cross-modal correspondence.

In summary, the present study used a cue-target paradigm with auditory stimuli presented before visual targets to investigate the influence of IOR on audiovisual cross-modal correspondence. The study comprised three experiments. Experiment 1 manipulated spatial cue validity and cross-modal correspondence consistency between auditory stimuli and visual targets to explore the relationship between IOR and audiovisual cross-modal correspondence. We hypothesized that the pitch-space cross-modal correspondence occurs at the perceptual level and would therefore interact with the IOR effect. Experiment 2 manipulated whether an irrelevant auditory stimulus was presented to investigate the relationship between IOR and sound presentation alone. Since cross-modal correspondence is a relative mapping requiring two tones with a high-low relationship to correspond with high and low positions (Chiou & Rich, 2012), a single pure tone would not produce cross-modal correspondence with visual target location. The purpose of Experiment 2 was twofold: first, to verify that in the present paradigm, the mere presentation of an auditory stimulus before a visual stimulus does not interact with IOR, whereas cross-modal correspondence between auditory and visual stimuli does; second, to rule out potential confounding effects of alerting. Based on previous research showing that alerting enhances stimulus perception in a top-down manner (Kusnir et al., 2011) while IOR affects perception in a bottom-up fashion (Berdica et al., 2017; Jia et al., 2019), we hypothesized that IOR would not interact with sound presentation, further supporting that the results of Experiment 1 reflect IOR's influence on audiovisual cross-modal correspondence. Experiment 3 manipulated SOA between cue and target to vary IOR magnitude and explore the underlying mechanism of IOR's modulation of audiovisual cross-modal correspondence. According to the principle of inverse effectiveness, we expected that increased SOA would reduce IOR, thereby decreasing the cross-modal correspondence effect at cued locations and weakening IOR's modulatory effect.

Experiment 1: The Relationship Between IOR and Audio-visual Cross-Modal Correspondence

Participants

Sample size was calculated using G*Power 3.1 software (Erdfelder et al., 2009; Faul et al., 2007). With Type I error probability set at 0.05, statistical power ($1 - \beta$) at 0.8, and medium effect size (f = 0.25) (Cohen, 1992), the required sample size was 24. We recruited 31 university students from Jiangsu Province (14 males, 17 females) aged 18–24 years. All participants were right-handed with normal hearing and normal or corrected-to-normal vision, and had no history of neurological or psychiatric disorders or brain injury. Participants received compensation upon completion of the experiment.

Apparatus and Materials

The experimental program was compiled using E-Prime 2.0 and run on a Dell 3020 MT computer. Stimuli were presented on a 23-inch Dell E2316Hf LCD monitor with a resolution of 1024 × 768 and a refresh rate of 60 Hz. Participants' heads were stabilized with a chinrest positioned 60 cm from the screen. The experiment was conducted in a dark, soundproof room.

All visual stimuli were drawn in black (RGB: 0, 0, 0) on a white background. Each trial displayed three square boxes (1.5° × 1.5°) arranged vertically on the screen, with one box at the center and the other two above and below it. Adjacent boxes were separated by 4.5° of visual angle. A central fixation point (1° × 1°) appeared within the central box. Cues were implemented by thickening the border of the box above or below the fixation point by 0.5°, while central cues were implemented by enlarging the central fixation point to 1.5° × 1.5°. The visual target was a disk (1° × 1°). Auditory stimuli were 250 Hz or 2500 Hz sine wave tones (50 ms duration), presented binaurally through Audio-Technica ATH-WS99 headphones at 65 dB.

Design and Procedure

Experiment 1 employed a 2 (spatial cue validity: valid vs. invalid) × 2 (cross-modal correspondence congruency: congruent vs. incongruent) within-subjects design, with reaction time and accuracy as dependent variables. In the cross-modal correspondence congruent condition, a high-pitched tone preceded a visual target at the high spatial position, and a low-pitched tone preceded a target at the low spatial position. In the incongruent condition, this mapping was reversed. The experiment consisted of five blocks of 53 trials each, including five catch trials, for a total of 265 trials. Participants completed 53 practice trials before the formal experiment, which lasted approximately 35 minutes.

The trial procedure is illustrated in [Figure 1: see original paper]. Each trial began with a fixation cross presented for 750 ms. The border of the box above or below the fixation point was then thickened for 50 ms as a peripheral cue, which was non-predictive of target location. After a 250-ms interval, the central fixation point was enlarged for 50 ms as a central cue. Following a 200-ms interval, the visual target appeared for 100 ms in either the upper or lower box. Participants were instructed to respond as quickly and accurately as possible upon detecting the visual target; no response was required for catch trials. If no response was made within 1000 ms, the trial automatically advanced to the next. No feedback regarding response accuracy was provided during the formal experiment.

Results and Analysis

Data from incorrect responses, no responses, and outliers (RTs < 100 ms or > 3 SD from the individual mean) were excluded from RT analysis, accounting for 1.09% of total data. As Experiment 1 involved a simple detection task, mean accuracy exceeded 98%; therefore, no further statistical analysis of accuracy was conducted.

A 2 (cue validity: valid vs. invalid) $\times$ 2 (cross-modal correspondence congruency: congruent vs. incongruent) repeated-measures ANOVA on RT revealed significant main effects of cue validity, $F(1, 30) = 122.26$, $p < 0.001$, $^{2}\mathrm{p} = 0.80$, with slower RTs on valid-cue trials (325 ms) than invalid-cue trials (288 ms), demonstrating the IOR effect. The main effect of cross-modal correspondence congruency was also significant, $F(1, 30) = 4.95$, $p = 0.034$, $^{2}\mathrm{p} = 0.14$, with faster RTs in the congruent condition (305 ms) than in the incongruent condition (308 ms), indicating a cross-modal correspondence effect. Critically, the interaction between cue validity and cross-modal correspondence congruency was significant, $F(1, 30) = 6.69$, $p = 0.015$, $^{2}\mathrm{p} = 0.18$, suggesting that IOR modulated audiovisual cross-modal correspondence.

Simple effects analysis revealed that when the cue was valid, RTs were significantly faster in the congruent condition (322 ms) than in the incongruent condition (327 ms), $t(30) = 3.26$, $p = 0.003$, Cohen's $d = 0.59$, 95% CI $= [-9.29, -2.13]$, demonstrating a cross-modal correspondence effect. When the cue was invalid, there was no significant difference between congruent (289 ms) and incongruent (288 ms) conditions, $t(30) < 1$, indicating no cross-modal correspondence effect. Further simple effects analysis showed that when cross-modal correspondence was congruent, RTs were significantly slower on valid-cue trials (322 ms) than invalid-cue trials (288 ms), $t(30) = 10.19$, $p < 0.001$, Cohen's $d = 1.83$, 95% CI $= [26.76, 40.19]$, demonstrating the IOR effect. Similarly, when cross-modal correspondence was incongruent, RTs were significantly slower on valid-cue trials (327 ms) than invalid-cue trials (288 ms), $t(30) = 10.76$, $p < 0.001$, Cohen's $d = 1.93$, 95% CI $= [31.79, 40.69]$, with IOR also present.

A paired-samples t-test on IOR magnitude (RT difference between valid and

invalid cues) revealed that the IOR effect was significantly smaller in the congruent condition (33 ms) than in the incongruent condition (39 ms), $t(30) = 2.59$, $p = 0.015$, Cohen's $d = 0.47$, 95% CI = [−10.31, −1.21], indicating that cross-modal correspondence partially offset the IOR effect.

The results of Experiment 1 support the hypothesis that both auditory pitch-visual spatial location cross-modal correspondence and the IOR effect occur at early perceptual stages, leading to an interaction between them. However, in the present study, auditory stimuli presented before visual targets could produce an alerting effect (Wiegand & Sander, 2019), which might differ in magnitude between cued and uncued locations (Botta et al., 2017) and potentially confound the results. To further support that the results of Experiment 1 were indeed due to IOR's modulation of cross-modal correspondence, Experiment 2 manipulated whether an auditory stimulus was presented to investigate the relationship between IOR and the mere presence of a sound. Experiment 2 used a single tone that was either present or absent before the visual stimulus. Since cross-modal correspondence requires two tones with a relative high-low relationship to correspond with spatial positions (Chiou & Rich, 2012), a single tone would not produce cross-modal correspondence with visual target location. The purposes of Experiment 2 were: (1) to verify that in the present paradigm, the mere presentation of an auditory stimulus before a visual stimulus does not interact with IOR, whereas cross-modal correspondence between auditory and visual stimuli does; and (2) to rule out potential confounding effects of alerting. Based on previous research showing that alerting enhances stimulus perception in a top-down manner (Kusnir et al., 2011) while IOR affects perception in a bottom-up fashion (Berdica et al., 2017; Jia et al., 2019), we hypothesized that IOR would not interact with sound presentation, further supporting that Experiment 1's results were due to IOR's influence on cross-modal correspondence.

Experiment 2: IOR and Sound Presentation

Participants

Sample size calculation using G*Power 3.1 (Erdfelder et al., 2009; Faul et al., 2007) with $\alpha = 0.05$, power = 0.8, and medium effect size (f* = 0.25) (Cohen, 1992) indicated a required sample of 24. We recruited 34 university students from Jiangsu Province (15 males, 19 females) aged 18–24 years. All were right-handed with normal hearing and normal or corrected-to-normal vision, with no history of neurological or psychiatric disorders or brain injury. Participants received compensation after completing the experiment.

Apparatus and Materials

The auditory stimulus in Experiment 2 was a 1600 Hz sine wave tone; all other apparatus and materials were identical to Experiment 1.

Design and Procedure

Experiment 2 employed a 2 (spatial cue validity: valid vs. invalid) × 2 (sound presentation: present vs. absent) within-subjects design, with RT and accuracy as dependent variables. The only difference from Experiment 1 was that the auditory stimulus was a single 1600 Hz pure tone that could be either present or absent before the visual stimulus. All other procedures and trial configurations remained identical to Experiment 1.

Results and Analysis

Data from incorrect responses, no responses, and outliers (RTs $<$ 100 ms or $>$ 3 SD from the individual mean) were excluded, accounting for 1.88% of total data. Mean accuracy exceeded 98%; therefore, no further accuracy analysis was conducted.

A 2 (cue validity: valid vs. invalid) × 2 (sound presentation: present vs. absent) repeated-measures ANOVA on RT revealed a significant main effect of cue validity, $F(1, 33) = 237.78$, $p < 0.001$, $^{2}\mathrm{p} = 0.88$, with slower RTs on valid-cue trials (313 ms) than invalid-cue trials (294 ms), demonstrating the IOR effect. The main effect of sound presentation was significant, $F(1, 33) = 82.34$, $p < 0.001$, $^{2}\mathrm{p} = 0.71$, with faster RTs when sound was present (283 ms) than absent (305 ms), indicating that the auditory stimulus facilitated visual target processing. Critically, the interaction between cue validity and sound presentation was not significant, $F(1, 33) < 1$, providing no evidence that IOR influenced the facilitatory effect of sound.

The results of Experiment 2 demonstrate that IOR does not interact with the mere presence of a sound; only when auditory and visual stimuli engage in cross-modal correspondence does an interaction with IOR occur. The facilitatory effect of auditory stimuli on visual target responses in Experiment 2 primarily reflects alerting, which did not interact with IOR. Consistent with previous research, alerting enhances stimulus perception in a top-down manner (Kusnir et al., 2011), whereas IOR affects perception in a bottom-up fashion (Berdica et al., 2017; Jia et al., 2019). Because these two effects operate via different pathways, they do not interact. Experiment 2 thus supports that the results of Experiment 1 were indeed due to IOR's influence on audiovisual cross-modal correspondence.

To further investigate the mechanism by which IOR modulates audiovisual cross-modal correspondence, Experiment 3 manipulated SOA between cue and target to vary IOR magnitude and examined how IOR magnitude influences cross-modal correspondence. If the principle of inverse effectiveness holds, then reduced IOR at longer SOAs should weaken the cross-modal correspondence effect at cued locations.

Experiment 3: Effects of IOR on Audiovisual Cross-Modal Correspondence at Different SOAs

Participants

Sample size calculation using G*Power 3.1 (Erdfelder et al., 2009; Faul et al., 2007) with $\alpha = 0.05$, power = 0.8, and medium effect size (f* = 0.25) (Cohen, 1992) indicated a required sample of 16. We recruited 37 university students from Jiangsu Province (9 males, 28 females). Three participants were excluded, leaving 34 valid participants (9 males, 25 females) aged 19–26 years. All were right-handed with normal hearing and normal or corrected-to-normal vision, with no history of neurological or psychiatric disorders or brain injury. Participants received compensation after completing the experiment.

Apparatus and Materials

Based on Experiment 1, Experiment 3 set the SOA between cue and target at two levels: 600 ms and 1300 ms. The 600 ms SOA was identical to Experiment 1, while the 1300 ms SOA was implemented by extending the interval between the peripheral and central cues. All other apparatus and materials were identical to Experiment 1.

Design and Procedure

Experiment 3 employed a 2 (spatial cue validity: valid vs. invalid) × 2 (cross-modal correspondence congruency: congruent vs. incongruent) × 2 (SOA: 600 ms vs. 1300 ms) within-subjects design, with RT and accuracy as dependent variables. The formal experiment consisted of 414 trials. Participants completed 35 practice trials before the formal experiment, which lasted approximately 50 minutes. All other procedures were identical to Experiment 1.

Results and Analysis

Data from incorrect responses, no responses, and outliers (RTs < 100 ms or > 3 SD from the individual mean) were excluded, accounting for 1.22% of total data. Mean accuracy exceeded 99%; therefore, no further accuracy analysis was conducted.

Reaction Time A 2 (cue validity: valid vs. invalid) × 2 (cross-modal correspondence congruency: congruent vs. incongruent) × 2 (SOA: 600 ms vs. 1300 ms) repeated-measures ANOVA on RT revealed significant main effects of cue validity, $F(1, 33) = 89.44$, $p < 0.001$, $^2p = 0.73$, with slower RTs on valid-cue trials (355 ms) than invalid-cue trials (336 ms), demonstrating the IOR effect. The main effect of cross-modal correspondence congruency was significant, $F(1, 33) = 9.57$, $p = 0.004$, $^2p = 0.23$, with faster RTs in the congruent condition (343 ms) than in the incongruent condition (348 ms), indicating a cross-modal correspondence effect. The main effect of SOA was not significant, $F(1, 33)$

< 1. The interaction between SOA and cue validity was significant, $F(1, 33) = 6.89$, $p = 0.013$, $^2\text{p} = 0.17$, indicating that SOA modulated the IOR effect. Simple effects analysis showed that at the 600 ms SOA, RTs were significantly slower on valid-cue trials (356 ms) than invalid-cue trials (334 ms), $t(33) = 8.34$, $p < 0.001$, Cohen's $d = 1.43$, 95% CI = [16.33, 26.86], demonstrating IOR. Similarly, at the 1300 ms SOA, RTs were significantly slower on valid-cue trials (354 ms) than invalid-cue trials (339 ms), $t(33) = 8.52$, $p < 0.001$, Cohen's $d = 1.46$, 95% CI = [12.13, 19.74], with IOR also present. The modulation of IOR by SOA was reflected in the finding that the IOR effect at 600 ms SOA (22 ms) was significantly larger than at 1300 ms SOA (16 ms), $t(33) = 2.63$, $p = 0.013$, Cohen's $d = 0.45$, 95% CI = [1.27, 10.05].

Critically, the three-way interaction among cue validity, cross-modal correspondence congruency, and SOA was significant, $F(1, 33) = 6.40$, $p = 0.016$, $^2\text{p} = 0.16$. At the 600 ms SOA, the interaction between cue validity and cross-modal correspondence congruency was significant, $F(1, 33) = 19.45$, $p < 0.001$, $^2\text{p} = 0.37$, indicating that IOR modulated cross-modal correspondence. Simple effects analysis showed that when the cue was valid, RTs were significantly faster in the congruent condition (350 ms) than in the incongruent condition (361 ms), $t(33) = 4.97$, $p < 0.001$, Cohen's $d = 0.85$, 95% CI = [−15.36, −6.43], demonstrating cross-modal correspondence. When the cue was invalid, there was no significant difference between congruent (334 ms) and incongruent (335 ms) conditions, $t(33) < 1$, indicating no cross-modal correspondence. At the 1300 ms SOA, the main effect of cross-modal correspondence congruency was significant, $F(1, 33) = 5.41$, $p = 0.026$, $^2\text{p} = 0.14$, with faster RTs in the congruent condition (344 ms) than in the incongruent condition (349 ms), indicating cross-modal correspondence. However, the interaction between cue validity and cross-modal correspondence congruency was not significant, $F < 1$, with cross-modal correspondence occurring at both cued and uncued locations. Notably, at cued locations, the cross-modal correspondence effect was statistically significant ($t(33) = 2.11$, $p = 0.042$, Cohen's $d = 0.36$, 95% CI = [−9.73, −0.19]), while at uncued locations it approached significance ($t(33) = 1.78$, $p = 0.084$, Cohen's $d = 0.31$, 95% CI = [−9.44, 0.63]), suggesting that at the long SOA, the cross-modal correspondence effect at cued locations was more robust.

Cross-Modal Correspondence Effect Cross-modal correspondence magnitude (RT difference between incongruent and congruent conditions) was calculated for each SOA and cue validity condition. A 2 (SOA: 600 ms vs. 1300 ms) × 2 (cue validity: valid vs. invalid) repeated-measures ANOVA revealed a significant main effect of cue validity, $F(1, 33) = 10.45$, $p = 0.003$, $^2\text{p} = 0.24$, with larger cross-modal correspondence effects at cued locations (8 ms) than at uncued locations (3 ms). The main effect of SOA was not significant, $F(1, 33) < 1$. The interaction between cue validity and SOA was significant, $F(1, 33) = 6.40$, $p = 0.016$, $^2\text{p} = 0.16$. Simple effects analysis showed that at cued locations, the cross-modal correspondence effect was significantly larger at 600 ms SOA (11 ms) than at 1300 ms SOA (5 ms), $t(33) = 2.20$, $p = 0.035$, Cohen's

$d = 0.38$, 95% CI = [0.44, 11.44]. At uncued locations, there was no significant difference between 600 ms (1 ms) and 1300 ms (4 ms) SOAs, $t(33) = 1.45$, $p = 0.156$. Further simple effects analysis revealed that at 600 ms SOA, the cross-modal correspondence effect was significantly larger at cued locations (11 ms) than at uncued locations (1 ms), $t(33) = 4.41$, $p < 0.001$, Cohen's $d = 0.76$, 95% CI = [5.35, 14.50]. At 1300 ms SOA, there was no significant difference between cued (5 ms) and uncued (4 ms) locations, $t(33) < 1$.

IOR Effect IOR magnitude was calculated for each SOA and cross-modal correspondence condition. A 2 (SOA: 600 ms vs. 1300 ms) × 2 (cross-modal correspondence congruency: congruent vs. incongruent) repeated-measures ANOVA revealed a significant main effect of SOA, $F(1, 33) = 6.89$, $p = 0.013$, $^2\text{p} = 0.17$, with larger IOR effects at 600 ms SOA (22 ms) than at 1300 ms SOA (16 ms), confirming that IOR decreased with longer SOA. The main effect of cross-modal correspondence congruency was significant, $F(1, 33) = 10.45$, $p = 0.003$, $^2\text{p} = 0.24$, with smaller IOR effects in the congruent condition (16 ms) than in the incongruent condition (21 ms). The interaction between SOA and cross-modal correspondence congruency was significant, $F(1, 33) = 6.40$, $p = 0.016$, $^2\text{p} = 0.16$. Simple effects analysis showed that at 600 ms SOA, the IOR effect was significantly smaller in the congruent condition (17 ms) than in the incongruent condition (27 ms), $t(33) = 4.41$, $p < 0.001$, Cohen's $d = 0.76$, 95% CI = [−14.50, −5.35], indicating that cross-modal correspondence partially offset IOR. At 1300 ms SOA, there was no significant difference in IOR effects between congruent (16 ms) and incongruent (16 ms) conditions, $t(33) < 1$. The absence of a difference at 1300 ms may reflect that the weakened cross-modal correspondence effect reduced its ability to counteract IOR.

Experiment 3 manipulated SOA between cue and target to vary IOR magnitude and investigate its influence on audiovisual cross-modal correspondence. Analysis of IOR magnitude confirmed that IOR decreased with longer SOA, consistent with previous research (Lupiáñez et al., 1997). The combined results showed that at 600 ms SOA, IOR interacted with cross-modal correspondence: a correspondence effect emerged at cued locations but not at uncued locations, replicating Experiment 1. As SOA increased to 1300 ms, the cross-modal correspondence effect at cued locations significantly decreased, and IOR's modulatory effect on cross-modal correspondence weakened, as evidenced by the non-significant interaction between cue validity and cross-modal correspondence congruency (no difference between cued and uncued locations). These results conform to the principle of inverse effectiveness (Meredith & Stein, 1983): at 1300 ms SOA, the reduced IOR effect increased the perceptual saliency of visual targets at cued locations compared with 600 ms SOA, and stronger visual input produced weaker cross-modal correspondence. Additionally, because IOR weakened, the difference in perceptual saliency between cued and uncued locations decreased, reducing IOR's modulatory effect and eliminating the difference in cross-modal correspondence between cued and uncued locations. However, because IOR remained present at 1300 ms SOA, perceptual saliency at cued

locations was still relatively lower, and statistical results showed that the cross-modal correspondence effect at cued locations was more robust than at uncued locations.

General Discussion

The present study investigated the effect of IOR on audiovisual cross-modal correspondence using a spatial cue-target paradigm with auditory stimuli presented before visual targets. Experiment 1 found an interaction between IOR and cross-modal correspondence: a cross-modal correspondence effect emerged only at cued locations, supporting the hypothesis that audiovisual cross-modal correspondence occurs at early perceptual stages and thus interacts with the IOR effect, which also occurs at the perceptual stage. Experiment 2 demonstrated that when the auditory stimulus was a single tone, IOR did not influence its facilitatory effect on visual targets, indicating that only when auditory and visual stimuli engage in cross-modal correspondence does an interaction with IOR occur. Furthermore, Experiment 2 showed that alerting effects induced by sound did not interact with IOR, further supporting that Experiment 1's results were due to IOR's influence on cross-modal correspondence. Experiment 3 manipulated SOA to vary IOR magnitude and found that as IOR weakened, the cross-modal correspondence effect at cued locations decreased, and IOR's modulatory effect on cross-modal correspondence diminished. These results support the applicability of the principle of inverse effectiveness to audiovisual cross-modal correspondence.

Experiment 1's finding of an interaction between IOR and cross-modal correspondence, according to the additive factors logic of reaction time (Sternberg, 1969), indicates that both factors operate at the same processing stage. Thus, the present results support that cross-modal correspondence between auditory pitch and visual spatial location occurs at the same perceptual processing stage as IOR, consistent with some previous research (Kovi ć et al., 2010; Maeda et al., 2004). Some studies have suggested that audiovisual cross-modal correspondence occurs at the semantic level (Gallace & Spence, 2006; Martino & Marks, 1999), proposing that visual and auditory information activate shared semantic codes that produce cross-modal correspondence. For example, researchers have used the spoken words "high" and "low" instead of actual tones and found cross-modal correspondence between speech and spatial location (Gallace & Spence, 2006). In addition to direct semantic relations, researchers have found cross-modal correspondence between pitch and brightness even when brightness stimuli were replaced with semantically related words such as "daytime" and "nighttime" (Martino & Marks, 1999). The present results, however, demonstrate that audiovisual cross-modal correspondence can occur at a purely perceptual level without requiring semantic mediation. This aligns with findings that populations who do not use "high" and "low" to describe pitch still show pitch-space cross-modal correspondence (Parkinson et al., 2012) and that pre-linguistic in-

fants exhibit audiovisual cross-modal correspondence (Dolscheid et al., 2014; Walker et al., 2010). Of course, the present findings do not deny that semantic coding may play a role in cross-modal correspondence. The cross-modal correspondence examined here involves two basic stimulus features (pitch and spatial location) that are naturally correlated in the environment (Spence, 2011); for instance, heavier animals tend to produce lower-frequency sounds (e.g., cows) and are less likely to occupy high positions (e.g., in the air). Additionally, the human larynx lowers when producing low pitches and rises when producing high pitches (Parkinson et al., 2012), which may lead to natural associations between pitch and spatial location. Nevertheless, the present results indicate that cross-modal correspondence between pitch and spatial location can occur through perceptual mechanisms alone.

Experiment 1 found that cross-modal correspondence occurred only at cued locations, similar to some findings in audiovisual integration research. Previous studies have found larger audiovisual integration effects at cued locations when IOR occurs (Tang et al., 2020), which has been explained by the principle of inverse effectiveness: weaker visual and auditory input produces stronger integration (Meredith & Stein, 1983). In the present study, when IOR occurred, the reduced perceptual saliency of targets at cued locations and the relatively increased saliency at uncued locations (Satel et al., 2013) may have resulted in cross-modal correspondence occurring only at cued locations. This suggests that the principle of inverse effectiveness may also apply to audiovisual cross-modal correspondence. Experiment 1 also found that the IOR effect was smaller under cross-modal correspondence congruent conditions, indicating that cross-modal correspondence partially offset IOR-induced suppression of early perceptual processing, consistent with findings from audiovisual integration research (Tang et al., 2019).

Experiment 3's results showed that as the IOR effect weakened, the cross-modal correspondence effect at cued locations decreased accordingly, and IOR's modulatory effect on cross-modal correspondence diminished, directly confirming the applicability of the principle of inverse effectiveness to audiovisual cross-modal correspondence. In research on multisensory response enhancement, the principle of inverse effectiveness has been observed not only in meaningless audiovisual integration (Senkowski et al., 2011) but also in speech perception under multisensory input (van de Rijt et al., 2019), where more difficult-to-perceive words produced greater multisensory response enhancement. The present findings extend the principle of inverse effectiveness to audiovisual cross-modal correspondence, broadening its applicability in the domain of multisensory response enhancement.

The study also found that audiovisual cross-modal correspondence influenced the IOR effect. Specifically, at 600 ms SOA, the IOR effect was significantly smaller under cross-modal correspondence congruent than incongruent conditions in both Experiments 1 and 3. This occurred because cross-modal correspondence between auditory and visual stimuli increased the perceptual saliency

of visual targets, partially offsetting IOR-induced reductions in saliency. As SOA increased, the difference in IOR effects between congruent and incongruent conditions disappeared, reflecting that the weakened cross-modal correspondence effect reduced its capacity to counteract IOR.

The present study is the first to investigate the relationship between IOR and audiovisual cross-modal correspondence. Therefore, it was essential to ensure that the observed interactions were indeed due to IOR's modulation of cross-modal correspondence. Experiment 2 manipulated sound presentation while holding other conditions constant to examine the effect of sound alone. Because cross-modal correspondence is a relative mapping requiring two tones with a high-low relationship to correspond with spatial positions (Chiou & Rich, 2012), a single pure tone would not produce cross-modal correspondence with visual target location. The results showed no interaction between IOR and sound presentation, verifying that only cross-modal correspondence between auditory and visual stimuli interacts with IOR. Additionally, Experiment 2 demonstrated that alerting effects from auditory stimuli did not interact with IOR. Previous research has shown that alerting effects enhance perception in a top-down manner (Kusnir et al., 2011) and interact with target perceptual saliency (Botta et al., 2017), and behavioral and neural evidence indicates that alerting interacts with spatial attention (Botta et al., 2014, 2017), with the arousal system associated with alerting showing compensatory mechanisms with the attention system (Fischer et al., 2008; Portas et al., 1998). However, Experiment 2's results suggest that in the present paradigm, alerting effects induced by exogenous cues lack interaction with IOR. This may be because previous studies compared interactions between alerting and suprathreshold, subthreshold, and threshold stimuli (Botta et al., 2017; Chica et al., 2016), whereas in the present study, visual targets at both cued and uncued locations were fully visible, and the difference in target perceptual saliency produced by IOR was insufficient to trigger alerting modulation. Alternatively, alerting and IOR may operate along independent neural pathways. Although both effects influence perceptual saliency (Botta et al., 2014; Prime & Jolicoeur, 2009) and are associated with activation of frontoparietal networks (Bourgeois et al., 2012; Kusnir et al., 2011), alerting enhances perception top-down by activating frontoparietal networks to amplify input stimulus intensity, whereas exogenous cue-induced IOR affects perception bottom-up by modulating input stimulus intensity and influencing early visual area projections to frontoparietal networks (Botta et al., 2014). The relationship between alerting and IOR requires further investigation at the neural level. The present study ruled out alerting as a confounding factor, supporting the conclusion that pitch-space cross-modal correspondence occurs at the perceptual level.

In summary, the present study found that IOR modulated audiovisual cross-modal correspondence: when IOR occurred, stable cross-modal correspondence effects emerged at cued locations but not at uncued locations. Alerting effects from auditory stimuli did not interact with IOR. As the IOR effect weakened, the cross-modal correspondence effect at cued locations decreased, and IOR's

modulatory effect on cross-modal correspondence diminished. These results support that cross-modal correspondence between auditory pitch and visual spatial location occurs at the perceptual level and conforms to the principle of inverse effectiveness.

References

Berdica, E., Gerdes, A. B. M., & Alpers, G. W. (2017). A comprehensive look at phobic fear in inhibition of return: Phobia-related spiders as cues and targets. *Journal of Behavior Therapy and Experimental Psychiatry, 54*, 158–165.

Botta, F., Lupiáñez, J., & Chica, A. B. (2014). When endogenous spatial attention improves conscious perception: Effects of alerting and bottom-up activation. *Consciousness and Cognition, 23*, 63–73.

Botta, F., Ródenas, E., & Chica, A. B. (2017). Target bottom-up strength determines the extent of attentional modulations on conscious perception. *Experimental Brain Research, 235*(7), 2109–2124.

Bourgeois, A., Chica, A. B., Migliaccio, R., Thiebaut de Schotten, M., & Bartolomeo, P. (2012). Cortical control of inhibition of return: Evidence from patients with inferior parietal damage and visual neglect. *Neuropsychologia, 50*(5), 800–809.

Brunel, L., Carvalho, P. F., & Goldstone, R. L. (2015). It does belong together: Cross-modal correspondences influence cross-modal integration during perceptual learning. *Frontiers in Psychology, 6*, 358.

Brunetti, R., Indraccolo, A., Del Gatto, C., Spence, C., & Santangelo, V. (2018). Are crossmodal correspondences relative or absolute? Sequential effects on speeded classification. *Attention, Perception & Psychophysics, 80*(2), 527–534.

Chica, A. B., Bayle, D. J., Botta, F., Bartolomeo, P., & Paz-Alonso, P. M. (2016). Interactions between phasic alerting and consciousness in the fronto-striatal network. *Scientific Reports, 6*, 31868.

Chica, A. B., Lasaponara, S., Chanes, L., Valero-Cabré, A., Doricchi, F., Lupiáñez, J., & Bartolomeo, P. (2011). Spatial attention and conscious perception: The role of endogenous and exogenous orienting. *Attention, Perception & Psychophysics, 73*(4), 1065–1081.

Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception, 41*(3), 339–353.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Dolscheid, S., Hunnius, S., Casasanto, D., & Majid, A. (2014). Prelinguistic infants are sensitive to space-pitch associations found across cultures. *Psychological Science, 25*(6), 1256–1265.

Erdfelder, E., Auer, T. S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Natural Science Journal of Harbin Normal University, 217*(3), 108–124.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of Vision, 10*(1), 1–12.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences.* Behavior Research Methods, 39*(2), 175–191.

Fischer, T., Langner, R., Birbaumer, N., & Brocke, B. (2008). Arousal and attention: Self-chosen stimulation optimizes cortical excitability and minimizes compensatory effort. *Journal of Cognitive Neuroscience, 20*(8), 1443–1453.

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research, 147*(3), 332–343.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics, 68*(7), 1191–1203.

Getz, L. M., & Kubovy, M. (2018). Questioning the automaticity of audiovisual correspondences. *Cognition, 175*, 101–108.

Hopfinger, J. B., & Mangun, G. R. (2001). Tracking the influence of reflexive attention on sensory and cognitive processing. *Cognitive, Affective & Behavioral Neuroscience, 1*(1), 56–65.

Jia, L., Wang, J., Zhang, K., Ma, H., & Sun, H. J. (2019). Do emotional faces affect inhibition of return? An ERP study. *Frontiers in Psychology, 10*(721), 1–8.

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Science, 4*(4), 138–147.

Ković, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition, 114*(1), 19–28.

Kusnir, F., Chica, A. B., Mitsumasu, M. A., & Bartolomeo, P. (2011). Phasic auditory alerting improves visual conscious perception. *Consciousness and Cognition, 20*(4), 1201–1210.

Lupiáñez, J., Milán, E. G., Tornay, F. J., Madrid, E., & Tudela, P. (1997). Does IOR occur in discrimination tasks? Yes, it does, but later. *Perception & Psychophysics, 59*(8), 1241–1254.

Maeda, F., Kanai, R., & Shimojo, S. (2004). Changing pitch induced visual motion illusion. *Current Biology, 14*(23), 990–991.

Maimon, N. B., Lamy, D., & Eitan, Z. (2020). Crossmodal correspondence between tonal hierarchy and visual brightness: Associating syntactic structure and perceptual dimensions across modalities. *Multisensory Research, 33*(8), 805–836.

Marks, L. E., Ben-Artzi, E., & Lakatos, S. (2003). Cross-modal interactions in auditory and visual discrimination. *International Journal of Psychophysiology, 50*(1–2), 125–145.

Martino, G., & Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis. *Perception, 28*(7), 903–923.

McCormick, K., Lacey, S., Stilla, R., Nygaard, L. C., & Sathian, K. (2018). Neural basis of the crossmodal correspondence between auditory pitch and visuospatial elevation. *Neuropsychologia, 112*, 19–30.

McCracken, H. S., Murphy, B. A., Glazebrook, C. M., Burkitt, J. J., Karellas, A. M., & Yielder, P. C. (2019). Audiovisual multisensory integration and evoked potentials in young adults with and without attention-deficit/hyperactivity disorder. *Frontiers in Human Neuroscience, 13*, 95.

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*(4608), 389–391.

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H. J., & Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *Journal of Neuroscience, 27*(42), 11431–11441.

Otto, T. U., Dassy, B., & Mamassian, P. (2013). Principles of multisensory behavior. *Journal of Neuroscience, 33*(17), 7463–7474.

Parise, C., & Spence, C. (2008). Synesthetic congruency modulates the temporal ventriloquism effect. *Neuroscience Letters, 442*(3), 257–261.

Parise, C., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: A study using the implicit association test. *Experimental Brain Research, 220*(3–4), 319–333.

Parkinson, C., Kohler, P. J., Sievers, B., & Wheatley, T. (2012). Associations between auditory pitch and visual elevation do not depend on language: Evidence from a remote population. *Perception, 41*(7), 854–861.

Peng, X., Chang, R., Li, Q., Wang, A., & Tang, X. (2019). Visually induced inhibition of return affects the audiovisual integration under different SOA conditions. *Acta Psychologica Sinica, 51*(7), 759–771.

Portas, C. M., Rees, G., Howseman, A. M., Josephs, O., Turner, R., & Frith, C. D. (1998). A specific role for the thalamus in mediating the interaction of attention and arousal in humans. *The Journal of Neuroscience, 18*(21), 8979–8989.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. G. Bowhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 531–556). Erlbaum.

Prime, D. J., & Jolicoeur, P. (2009). On the relationship between occipital cortex activity and inhibition of return. *Psychophysiology, 46*(6), 1278–1287.

Rach, S., Diederich, A., & Colonius, H. (2011). On quantifying multisensory interaction effects in reaction time and detection rate. *Psychological Research, 75*(2), 77–94.

Redden, R. S., Maclnnes, W. J., & Klein, R. M. (2021). Inhibition of return: An information processing theory of its natures and significance. *Cortex, 135*, 30–48.

Santangelo, V., Ho, C., & Spence, C. (2008). Capturing spatial attention with multisensory cues. *Psychonomic Bulletin & Review, 15*(2), 398–403.

Satel, J., Hilchey, M. D., Wang, Z. G., Story, R., & Klein, R. M. (2013). The effects of ignored versus foveated cues upon inhibition of return: An event-related potential study. *Attention, Perception, & Psychophysics, 75*(1), 29–40.

Senkowski, D., Saint-Amour, D., Höfle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *Neuroimage, 56*(4), 2200–2208.

Slagter, H. A., Prinssen, S., Reteig, L. C., & Mazaheri, A. (2016). Facilitation and inhibition in attention: Functional dissociation of pre-stimulus alpha activity, P1, and N1 components. *Neuroimage, 125*(6), 25–35.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics, 73*(4), 971–995.

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences, 1296*(1), 31–49.

Spence, C. (2019). On the relative nature of (pitch-based) crossmodal correspondences. *Multisensory Research, 32*(3), 235–265.

Starke, J., Ball, F., Heinze, H. J., & Noesselt, T. (2017). The spatio-temporal profile of multisensory integration. *The European Journal of Neuroscience, 51*(5), 1210–1223.

Stein, B. E., & Meredith, M. A. (1993). The merging of the senses. *Journal of Cognitive Neuroscience, 5*(3), 373–374.

Stein, B. E., Meredith, M. A., Huneycutt, W. S., & McDade, L. (1989). Behavioral indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli. *Journal of Cognitive Neuroscience, 1*(1), 12–24.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9*(4), 255–266.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30*, 276–315.

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.

Tang, X., Gao, Y., Yang, W., Ren, Y., Wu, J., Zhang, M., & Wu, Q. (2019). Bimodal-divided attention attenuates visually induced inhibition of return with audiovisual targets. *Experimental Brain Research, 237*(4), 1093–1107.

Tang, X., Sun, J., & Peng, X. (2020). The effect of bimodal divided attention on inhibition of return with audiovisual targets. *Acta Psychologica Sinica, 52*(3), 257–268.

Tang, X., Wu, J., & Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. *Neuroscience and Biobehavioral Reviews, 61*, 208–224.

van de Rijt, L. P. H., Roye, A., Mylanus, E. A. M., van Opstal, A. J., & van Wanrooij, M. M. (2019). The principle of inverse effectiveness in audiovisual speech perception. *Frontiers in Human Neuroscience, 13*, 335.

van der Stoep, N., Spence, C., Nijboer, T. C., & van der Stigchel, S. (2015). On the relative contributions of multisensory integration and crossmodal exogenous spatial attention to multisensory response enhancement. *Acta Psychologica, 162*, 20–28.

van der Stoep, N., van der Stigchel, S., & Nijboer, T. C. W. (2015). Exogenous spatial attention decreases audiovisual integration. *Attention Perception & Psychophysics, 77*(1), 464–482.

van der Stoep, N., van der Stigchel, S., Nijboer, T. C. W., & Spence, C. (2016). Visually induced inhibition of return affects the integration of auditory and visual information. *Perception, 46*(1), 6–17.

Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of connotative meaning: Visual angularity is hard, high-pitched, and bright. *Attention, Perception & Psychophysics, 74*(8), 1792–1809.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2010). Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. *Psychological Science, 21*(1), 21–25.

Wiegand, I., & Sander, M. C. (2019). Cue-related processing accounts for age differences in phasic alerting. *Neurobiology of Aging, 79*, 93–100.

Zeljko, M., Kritikos, A., & Grove, P. M. (2019). Lightness/pitch and elevation/pitch crossmodal correspondences are low-level sensory effects. *Attention, Perception & Psychophysics, 81*(5), 1609–1623.

Note: Figure translations are in progress. See original paper for figures.