Effects of Inhibition of Return on Auditory-Visual Cross-Modal Correspondence

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

Date: 2023-02-23T00:00:00+00:00

Abstract

Audiovisual crossmodal correspondence has been widely documented across various types of visual and auditory stimuli, yet the stage at which it occurs remains unclear. The present study employed a cue-target paradigm to investigate the influence of inhibition of return (IOR) on audiovisual crossmodal correspondence. Experiment 1 manipulated both the spatial congruency between cue and target and the crossmodal correspondence congruency between auditory pitch and visual target location. The results revealed an interaction between IOR and audiovisual crossmodal correspondence: a stable crossmodal correspondence effect was observed at the cued location, whereas this effect disappeared at the uncued location. Experiment 2 manipulated the presence versus absence of an irrelevant auditory stimulus and found no interaction between IOR and the mere occurrence of sound, thereby ruling out alerting effects as a confounding factor. Experiment 3 extended the stimulus onset asynchrony (SOA) between cue and target and demonstrated that as the IOR effect diminished, the audiovisual crossmodal correspondence effect at the cued location also weakened, and the modulatory influence of IOR on crossmodal correspondence decreased. These findings indicate that only when crossmodal correspondence occurs between auditory stimuli and visual spatial locations can it interact with the IOR effect, which also operates at the perceptual level, thereby supporting the notion that audiovisual crossmodal correspondence arises 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

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 the spatial consistency between cues and targets as well as 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 and found no interaction between IOR and the mere presence of sound, ruling out potential confounds from alerting effects. Experiment 3 extended the stimulus onset asynchrony (SOA) between cue and target and found 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 engage in cross-modal correspondence with visual spatial positions does an interaction occur with IOR, which also occurs at the perceptual level, supporting the view that audiovisual cross-modal correspondence arises during perceptual processing. The results also support the principle of inverse effectiveness for audiovisual cross-modal correspondence.

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

1. Introduction

Humans rely on multiple sensory modalities to perceive their environment, and the combination of signals from different modalities can facilitate 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., 2020). Multisensory integration refers to the process by which individuals combine information from different sensory modalities to form coherent and meaningful representations when such information possesses spatiotemporal proximity (彭姓 et al., 2019; Tang et al., 2016). The integration of visual and auditory inputs, termed audiovisual integration, produces a redundant effect that facilitates detection and discrimination of bimodal stimuli compared with unimodal visual or auditory stimuli alone (彭姓 et al., 2019; 唐晓雨 et al., 2020; Stein & Stanford, 2008; Talsma & Woldorff, 2005). For example, visual and auditory stimuli presented simultaneously at the same location provide identical spatial and temporal information, thereby enhancing an individual' s response capability (Spence, 2013).

Beyond influencing responses in a redundant manner, visual and auditory stimuli can also affect behavior through non-redundant mapping, a phenomenon termed audiovisual cross-modal correspondence (Spence, 2011, 2019). A common instance 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 targets presented at high spatial locations than at low locations, and vice versa (Chiou & Rich, 2012; Evans & Treisman, 2010; McCormick et al., 2018; Spence, 2019; Zeljko et al., 2019). Additionally, cross-modal correspondences have been documented 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), 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 correspondence between pitch and spatial location, there is no absolute mapping between a specific frequency and a particular height, but rather a mapping between the relatively higher or lower of two pitches and the relatively higher or lower of two positions.

Current research presents divergent views regarding the stage at which audiovisual cross-modal correspondence occurs. The first perspective posits that it arises at the perceptual level, enhancing the perceptual saliency of stimuli (Evans & Treisman, 2010). Studies have shown that when participants view ambiguously moving gratings, ascending pitch biases perception toward upward motion while descending pitch biases perception toward downward motion (Maeda et al., 2004). ERP research has demonstrated that when a preceding sound corresponds to the current visual symbol, the amplitude of the early perceptual component N1 is larger than when correspondence is absent, supporting the notion that cross-modal correspondence reflects perceptual enhancement (Ković et al., 2010). The second perspective suggests that cross-modal correspondence occurs at the semantic level (Spence, 2011). Researchers have noted that across nearly all languages, people describe different frequencies using "high" and "low," terms that also correspond to spatial positions and spatial frequencies. This shared semantic coding across modalities may underlie cross-modal correspondence (Walker, 2012). Supporting this semantic-level account, studies using the spoken words "high" and "low" instead of actual pitch frequencies have observed cross-modal correspondence between speech and spatial location (Gallace & Spence, 2006). The third perspective proposes that cross-modal correspondence occurs at a late decision-making level, wherein correspondence between modalities lowers the response criterion for target detection (Spence, 2011). Using signal detection theory, researchers have found that cross-modal correspondence does not affect perceptual sensitivity (d') but increases false alarm rates, suggesting that if cross-modal correspondence occurred at the perceptual level, enhanced response capability would not be accompanied by increased false alarms, thus implicating decision-level rather than perceptual-level processing (Marks et al., 2003). In summary, due to variations in stimulus materials and measurement indices, the stage at which audiovisual cross-modal correspondence occurs remains unresolved (Spence, 2011).

Both multisensory integration and correspondence can enhance target perceptual saliency (Evans & Treisman, 2010; Ković et al., 2010; Tang et al., 2019) and facilitate behavioral responses, while inhibition of return (IOR) in the attentional system also influences target perception (唐晓雨 et al., 2020; Tang et al., 2019). IOR refers to the phenomenon in cue-target paradigms where, when the stimulus onset asynchrony (SOA) between cue and target exceeds approximately 300 ms, response times to targets appearing at cued locations are slower than those to targets at uncued locations (Posner & Cohen, 1984). IOR prevents repeated searching of the same location and improves visual search efficiency (Redden et al., 2021). Although different theoretical explanations exist regarding IOR' s mechanism, it is widely accepted 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 P1 and N1 components evoked by stimuli at cued locations have lower amplitudes than those at uncued locations (Hopfinger & Mangun, 2001; Prime & Jolicoeur, 2009), supporting the view that IOR occurs at early perceptual stages. Previous research has examined interactions between IOR and multisensory stimuli, primarily focusing on audiovisual integration (彭姓 et al., 2019; Tang et al., 2019; van der Stoep, van der Stigchel, et al., 2015). In cue-target paradigms with audiovisual targets, IOR has been found to modulate audiovisual integration, with some studies reporting smaller integration effects at cued locations (彭姓 et al., 2019; Tang et al., 2019; van der Stoep et al., 2016) and others reporting larger effects (唐晓雨 et al., 2020). These divergent findings may relate to different SOA settings across experiments (唐晓雨 et al., 2020), but existing research consistently indicates that audiovisual integration occurs at the perceptual stage (Tang et al., 2019) and is therefore modulated by IOR, which also occurs at perceptual processing stages (彭姓 et al., 2019; 唐晓雨 et al., 2020).

Although audiovisual integration and cross-modal correspondence enhance multisensory responses through different mechanisms—the former via redundant information from spatiotemporal proximity (Noesselt et al., 2007; Santangelo et al., 2008) and the latter via mapping between different dimensional information (Chiou & Rich, 2012; McCormick et al., 2018)—if cross-modal correspondence occurs at the perceptual stage, then based on previous findings of interactions between multisensory stimuli and attentional cueing effects including IOR (唐晓雨 et al., 2020; Botta et al., 2017; Chica et al., 2011; Tang et al., 2019) and the logic of additive factors methodology (Sternberg, 1969), the reduced perceptual saliency of targets at cued locations during IOR should affect cross-modal correspondence. Conversely, if cross-modal correspondence occurs at semantic or decision levels, IOR should not influence it. Therefore, this study combined spatial cueing paradigms with cross-modal correspondence paradigms to investigate the relationship between IOR and audiovisual cross-modal correspondence. In previous cross-modal correspondence research, visual and auditory stimuli were presented synchronously (Brunel et al., 2015; Gallace & Spence, 2006; Getz & Kubovy, 2018), potentially confounding measurements with both audiovisual integration and cross-modal correspondence. Given that spatiotemporal proximity is necessary for audiovisual integration (Spence, 2011) and that integration effects disappear when visual-auditory stimulus intervals exceed 100 ms (van der Stoep, Spence, et al., 2015), the present study presented visual targets 200 ms after auditory stimulus offset and delivered auditory stimuli binaurally to minimize audiovisual integration effects.

Additionally, auditory stimuli preceding visual targets can produce alerting effects (Wiegand & Sander, 2019) that enhance perceptual capability for visual stimuli (Kusnir et al., 2011) and interact with target perceptual saliency (Botta et al., 2017). Such alerting effects may interact with exogenous cue-induced IOR, creating differential alerting across conditions. As this is the first study to examine the relationship between IOR and audiovisual cross-modal correspondence, controlling for potential confounding factors is essential. Therefore, we designed a control experiment to exclude possible interactions between alerting effects and IOR that could confound the results. Finally, to further investigate the mechanism underlying the interaction between IOR and cross-modal correspondence, we manipulated SOA between cue and target to vary IOR magnitude (Lupiáñez et al., 1997) and examined how IOR modulates cross-modal correspondence.

In summary, this study employed a cue-target paradigm with auditory stimuli preceding visual targets to investigate IOR' s influence on audiovisual cross-modal correspondence. The study comprised three experiments. Experiment 1 manipulated spatial cue validity and cross-modal correspondence consistency to explore the relationship between IOR and audiovisual cross-modal correspondence. We hypothesized that pitch-location cross-modal correspondence occurs at the perceptual level and would therefore be influenced by IOR at the same processing stage, resulting in an interaction between the two effects. Experiment 2 manipulated whether an auditory stimulus was presented to examine the relationship between IOR and mere sound presentation. Since cross-modal correspondence between pitch and spatial location requires two tones with a relative high-low relationship (Chiou & Rich, 2012), presenting a single tone would not produce cross-modal correspondence with visual target location. Experiment 2 aimed to verify that mere sound presentation does not interact with IOR, whereas cross-modal correspondence does, and to exclude potential confounding from alerting effects. Based on previous research showing that alerting effects enhance perception in a top-down manner (Kusnir et al., 2011) while IOR affects perception bottom-up (Berdica et al., 2017; Jia et al., 2019), we hypothesized that IOR would not interact with sound presence, further supporting that Experiment 1' s results reflect IOR' s influence on cross-modal correspondence. Experiment 3 manipulated SOA to vary IOR magnitude and investigate its modulatory mechanism. According to the principle of inverse effectiveness in multisensory response enhancement (Meredith & Stein, 1983; van der Stoep et al., 2016), we expected that increased SOA would reduce IOR (Lupiáñez et al., 1997), thereby decreasing cross-modal correspondence at cued locations and weakening IOR' s modulatory effect on cross-modal correspondence.

2.1.1 Participants

Sample size was calculated using G*Power 3.1 software (Erdfelder et al., 2009; Faul et al., 2007) with an alpha error probability of 0.05, power (1 - β error probability) of 0.8, and medium effect size (f = 0.25) (Cohen, 1992), yielding a required sample of 24 participants. We recruited 31 university students from Jiangsu Province (14 males, 17 females; age 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.

2.1.2 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 LCD monitor (Dell E2316Hf) 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 throughout the experiment, which was conducted in a dark, soundproof environment.

All visual stimuli were drawn in black (RGB: 0, 0, 0) lines 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 rectangle 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 50-ms sine tones of 250 Hz or 2500 Hz, presented binaurally through Audio-Technica headphones (ATH-WS99) at 65 dB.

2.1.3 Design and Procedure

Experiment 1 employed a 2 (spatial cue validity: valid vs. invalid) × 2 (cross-modal correspondence consistency: congruent vs. incongruent) within-subjects design, with reaction time and accuracy as dependent variables. The congruent condition presented high-pitch tones followed by high-location visual targets and low-pitch tones followed by low-location targets, while the incongruent condition reversed these mappings. The formal experiment comprised 5 blocks of 53 trials each, including 5 catch trials, for a total of 265 trials. Participants completed 35 practice trials before the formal experiment, which lasted approximately 40 minutes.

The trial procedure for Experiment 1 is illustrated in Figure 1 [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 then thickened for 50 ms as a peripheral cue, which was non-predictive of target location. After a 250-ms interval, the fixation point enlarged as a central cue for 50 ms. Central cues are commonly used in spatial IOR research to facilitate stable IOR effects (Prime et al., 2006). Following the central cue, an auditory stimulus was presented for 50 ms. After a 200-ms interval, a 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; catch trials required no response. If no response was made within 1000 ms, the next trial began automatically. No feedback on response accuracy was provided during the experiment, except during practice trials.

2.2 Results and Analysis

Trials with incorrect responses, no responses, or outlier reaction times (less than 100 ms or exceeding $±$3 standard deviations from the mean) were excluded from analysis, accounting for 1.09% of total data. Given that Experiment 1 involved a simple detection task with mean accuracy exceeding 98%, no further statistical analysis was conducted on accuracy.

A 2 (cue validity: valid vs. invalid) × 2 (cross-modal correspondence consistency: congruent vs. incongruent) repeated-measures ANOVA on reaction times revealed significant main effects of cue validity, $F(1, 30) = 122.26$, $p < 0.001$, $^2p = 0.80$, with slower responses on valid-cue trials (325 ms) than invalid-cue trials (288 ms), demonstrating an IOR effect. The main effect of cross-modal correspondence consistency was also significant, $F(1, 30) = 4.95$, $p = 0.034$, $^2p = 0.14$, with faster responses in congruent conditions (305 ms) than incongruent conditions (308 ms), indicating a cross-modal correspondence effect.

Critically, the interaction between cue validity and cross-modal correspondence consistency was significant, $F(1, 30) = 6.69$, $p = 0.015$, $^2p = 0.18$, showing that IOR modulated audiovisual cross-modal correspondence.

Simple effects analysis revealed that when the cue was valid, reaction times 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], indicating a cross-modal correspondence effect. When the cue was invalid, no significant difference emerged between congruent (289 ms) and incongruent (288 ms) conditions, $t(30) < 1$, indicating no cross-modal correspondence effect. Analysis from the other dimension showed that when cross-modal correspondence was congruent, valid-cue trials yielded significantly slower responses (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 IOR. This pattern also held for incongruent trials, with slower responses 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]. A paired-samples t-test on IOR magnitude (valid-cue RT minus invalid-cue RT) showed 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 IOR.

The results of Experiment 1 demonstrated slower responses on valid-cue than invalid-cue trials, confirming the IOR effect and indicating that presenting auditory stimuli after the central cue did not disrupt IOR. Additionally, faster responses in congruent than incongruent trials confirmed cross-modal correspondence between auditory pitch and visual location. Importantly, IOR interacted with cross-modal correspondence: the correspondence effect emerged only at cued locations, not at uncued locations, indicating that early perceptual-stage IOR modulated cross-modal correspondence.

Previous audiovisual integration research has found larger integration effects at cued locations during IOR (唐晓雨 et al., 2020), explained by the principle of inverse effectiveness—where weaker sensory input produces stronger integration (Meredith & Stein, 1983). In the present study, IOR reduced perceptual saliency at cued locations while relatively increasing it at uncued locations (Satel et al., 2013), potentially causing stable cross-modal correspondence only at cued locations. This suggests that the principle of inverse effectiveness may also apply to cross-modal correspondence, a point discussed further below. Experiment 1 also found that the IOR effect was smaller under congruent conditions, indicating that cross-modal correspondence partially counteracted IOR' s early perceptual inhibition, consistent with audiovisual integration findings (Tang et al., 2019).

Experiment 1' s results support the view that both pitch-location cross-modal correspondence and IOR occur at early perceptual stages, leading to their interaction. However, the auditory stimulus preceding the visual target may produce alerting effects (Wiegand & Sander, 2019) that could differ between cued and uncued locations (Botta et al., 2017), potentially confounding the results. To strengthen the conclusion that Experiment 1' s results reflect IOR' s modulation of cross-modal correspondence rather than alerting effects, Experiment 2 manipulated whether an auditory stimulus was presented to examine the relationship between IOR and mere sound presentation. Since cross-modal correspondence requires two tones with a relative relationship (Chiou & Rich, 2012), a single pure tone would not produce correspondence with visual target location. Experiment 2 aimed to verify that mere sound presentation does not interact with IOR, whereas cross-modal correspondence does, and to exclude potential alerting confounds.

3.1.1 Participants

Sample size calculation using G*Power 3.1 with the same parameters ($\alpha = 0.05$, power = 0.8, f = 0.25) indicated a required sample of 24 participants. We recruited 34 university students from Jiangsu Province (15 males, 19 females; age 18-24 years) with the same inclusion criteria as Experiment 1.

3.1.2 Design and Procedure

Experiment 2 employed a 2 (spatial cue validity: valid vs. invalid) × 2 (sound presentation: present vs. absent) within-subjects design, with reaction time 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 presented before the visual target or omitted entirely. All other procedures and trial structures remained identical to Experiment 1.

3.2 Results and Analysis

Trials with errors, no responses, or outlier reaction times (less than 100 ms or exceeding \$±\$3 SD) were excluded, accounting for 1.88% of data. Mean accuracy exceeded 98%, so no further accuracy analysis was conducted.

A 2 (cue validity: valid vs. invalid) × 2 (sound presentation: present vs. absent) repeated-measures ANOVA on reaction times revealed a significant main effect of cue validity, $F(1, 33) = 237.78$, $p < 0.001$, ^{2}p = 0.88, with slower responses on valid-cue trials (313 ms) than invalid-cue trials (294 ms), demonstrating IOR. The main effect of sound presentation was also significant, $F(1, 33) = 82.34$, $p < 0.001$, ^{2}p = 0.71, with faster responses when sound was present (283 ms) than absent (305 ms), indicating that auditory stimuli facilitated visual target detection. 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.

Experiment 2 demonstrated slower responses on valid-cue than invalid-cue trials, confirming IOR, and showed that auditory stimuli presented 200 ms before visual targets facilitated responses. However, no interaction emerged between cue validity and sound presentation. These results indicate that mere sound presentation does not interact with IOR; only when auditory stimuli engage in cross-modal correspondence with visual targets does an interaction occur. The facilitatory effect in Experiment 2 primarily reflected alerting effects, which did not interact with IOR. Consistent with previous research showing that alerting enhances perception top-down (Kusnir et al., 2011) while IOR affects perception bottom-up (Berdica et al., 2017; Jia et al., 2019), the two effects operate via different pathways and thus do not interact. Experiment 2 supports the conclusion that Experiment 1' s results indeed reflect IOR' s influence on cross-modal correspondence.

To further investigate the modulatory mechanism of IOR on cross-modal correspondence, Experiment 3 manipulated SOA to vary IOR magnitude and examined its impact on cross-modal correspondence. If the principle of inverse effectiveness holds, increased SOA should reduce IOR (Lupiáñez et al., 1997), thereby decreasing cross-modal correspondence at cued locations and weakening IOR' s modulatory effect.

4.1.1 Participants

Sample size calculation ($\alpha = 0.05$, power = 0.8, f = 0.25) indicated a required sample of 16 participants. We recruited 37 university students from Jiangsu Province (9 males, 28 females). Three participants were excluded, leaving 34 valid participants (9 males, 25 females; age 19-26 years) with the same inclusion criteria as previous experiments.

4.1.2 Apparatus and Materials

Experiment 3 built upon Experiment 1 by setting two SOA levels between cue and target: 600 ms (identical to Experiment 1) and 1300 ms (achieved by extending the interval between peripheral and central cues). All other apparatus and materials remained identical to Experiment 1.

4.1.3 Design and Procedure

Experiment 3 employed a 2 (spatial cue validity: valid vs. invalid) × 2 (cross-modal correspondence consistency: congruent vs. incongruent) × 2 (SOA: 600 ms vs. 1300 ms) within-subjects design, with reaction time and accuracy as dependent variables. The formal experiment comprised 6 blocks of 69 trials each, including 5 catch trials, for a total of 414 trials. Participants completed 35 practice trials before the formal experiment, which lasted approximately 50 minutes. All other procedures remained identical to Experiment 1.

4.2 Results and Analysis

Trials with errors, no responses, or outlier reaction times (less than 100 ms or exceeding $±$3 SD) were excluded, accounting for 1.22% of data. Mean accuracy exceeded 99%, so no further accuracy analysis was conducted.

4.2.1 Reaction Time A 2 (cue validity: valid vs. invalid) × 2 (cross-modal correspondence consistency: congruent vs. incongruent) × 2 (SOA: 600 ms vs. 1300 ms) repeated-measures ANOVA on reaction times (see Table 1) revealed significant main effects of cue validity, $F(1, 33) = 89.44$, $p < 0.001$, ^{2}p = 0.73, with slower responses on valid-cue trials (355 ms) than invalid-cue trials (336 ms), confirming IOR, and cross-modal correspondence consistency, $F(1, 33) = 9.57$, $p = 0.004$, ^{2}p = 0.23, with faster responses in congruent (343 ms) than incongruent (348 ms) conditions. 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}p = 0.17, indicating that SOA modulated IOR. Simple effects analysis showed that at 600-ms SOA, valid-cue trials yielded slower responses (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 1300-ms SOA, valid-cue trials were slower (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], also demonstrating IOR. The modulation of IOR by SOA was evident in that the IOR effect was significantly larger at 600-ms SOA (22 ms) 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], showing that IOR decreased with longer SOA.

Critically, the three-way interaction among cue validity, cross-modal correspondence consistency, and SOA was significant, $F(1, 33) = 6.40$, $p = 0.016$, ^{2}p = 0.16. At 600-ms SOA, the interaction between cue validity and cross-modal correspondence consistency was significant, $F(1, 33) = 19.45$, $p < 0.001$, ^{2}p = 0.37, indicating IOR modulation of cross-modal correspondence. Simple effects analysis revealed that when the cue was valid, congruent trials yielded faster responses (350 ms) than incongruent trials (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, no significant difference emerged between congruent (334 ms) and incongruent (335 ms) conditions, $t(33) < 1$, indicating no cross-modal correspondence. At 1300-ms SOA, the main effect of cross-modal correspondence consistency was significant, $F(1, 33) = 5.41$, $p = 0.026$, ^{2}p = 0.14, with faster responses in congruent (344 ms) than incongruent (349 ms) conditions, showing cross-modal correspondence. However, the interaction between cue validity and cross-modal correspondence consistency was not significant, $F < 1$, with cross-modal correspondence appearing in both valid- and invalid-cue conditions. Notably, at this longer SOA, the cross-modal correspondence effect was statistically significant at cued locations ($t(33) = 2.11$, $p = 0.042$, Cohen' s d = 0.36, 95% CI [-9.73, -0.19]) and marginally significant at uncued locations ($t(33) = 1.78$, $p = 0.084$, Cohen' s d = 0.31, 95% CI [-9.44, 0.63]), suggesting more stable cross-modal correspondence at cued locations even with long SOA.

4.2.2 Cross-Modal Correspondence Effect Cross-modal correspondence magnitude (incongruent minus congruent mean RT) was calculated for each

SOA and cue validity condition and submitted to a 2 (SOA: 600 ms vs. 1300 ms) × 2 (cue validity: valid vs. invalid) repeated-measures ANOVA.

As shown in Figure 4 [Figure 4: see original paper], the main effect of cue validity was significant, $F(1, 33) = 10.45$, $p = 0.003$, $^2p = 0.24$, with larger correspondence effects at cued locations (8 ms) than 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, $F(1, 33) = 6.40$, $p = 0.016$, $^2p = 0.16$. Simple effects analysis revealed that at cued locations, the 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, no significant difference emerged between 600-ms (1 ms) and 1300-ms (4 ms) SOA conditions, $t(33) = 1.45$, $p = 0.156$. Conversely, at 600-ms SOA, the correspondence effect was significantly larger at cued locations (11 ms) than uncued locations (1 ms), $t(33) = 4.41$, $p < 0.001$, Cohen' s d = 0.76, 95% CI [5.35, 14.50], whereas at 1300-ms SOA, no significant difference emerged between cued (5 ms) and uncued (4 ms) locations, $t(33) < 1$.

4.2.3 IOR Effect IOR magnitude was calculated for each SOA and cross-modal correspondence condition and submitted to a 2 (SOA: 600 ms vs. 1300 ms) × 2 (cross-modal correspondence consistency: congruent vs. incongruent) repeated-measures ANOVA.

Results showed a significant main effect of SOA, $F(1, 33) = 6.89$, $p = 0.013$, $^2p = 0.17$, 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 consistency was also significant, $F(1, 33) = 10.45$, $p = 0.003$, $^2p = 0.24$, with smaller IOR effects in congruent (16 ms) than incongruent (21 ms) conditions. The interaction between SOA and cross-modal correspondence consistency was significant, $F(1, 33) = 6.40$, $p = 0.016$, $^2p = 0.16$. Simple effects analysis revealed that at 600-ms SOA, the IOR effect was significantly smaller in congruent (17 ms) than incongruent (27 ms) conditions, $t(33) = 4.41$, $p < 0.001$, Cohen' s d = 0.76, 95% CI [-14.50, -5.35], indicating that cross-modal correspondence partially offset IOR. At 1300-ms SOA, no significant difference emerged between congruent (16 ms) and incongruent (16 ms) conditions, $t(33) < 1$.

Experiment 3 manipulated SOA to vary IOR magnitude and examine its impact on cross-modal correspondence. Analysis of IOR magnitude confirmed that IOR decreased with longer SOA, consistent with previous research (Lupiáñez et al., 1997). At 600-ms SOA, IOR interacted with cross-modal correspondence: the correspondence effect emerged at cued but not uncued locations, replicating Experiment 1. At 1300-ms SOA, the correspondence effect at cued locations decreased significantly compared to 600-ms SOA, and IOR' s modulatory effect weakened, as evidenced by the non-significant interaction between cue validity and cross-modal correspondence consistency (no difference between cued and uncued locations). These results support the principle of inverse effectiveness

(Meredith & Stein, 1983): at 1300-ms SOA, reduced IOR enhanced perceptual saliency at cued locations, and stronger visual input produced weaker cross-modal correspondence. Additionally, reduced IOR diminished the difference in perceptual saliency between cued and uncued locations, weakening IOR's modulatory effect and eliminating differences in cross-modal correspondence between locations. Nevertheless, because IOR persisted at 1300-ms SOA, perceptual saliency remained relatively lower at cued locations, resulting in more stable cross-modal correspondence at cued than uncued locations. Furthermore, the absence of a significant difference in IOR magnitude between congruent and incongruent conditions at 1300-ms SOA likely reflects weakened cross-modal correspondence reducing its capacity to counteract IOR.

5. Discussion

Using a spatial cue-target paradigm with auditory stimuli preceding visual targets, this study manipulated spatial cue validity and audiovisual cross-modal correspondence consistency to examine IOR' s influence on cross-modal correspondence. Experiment 1 revealed an interaction between IOR and cross-modal correspondence, with correspondence effects emerging only at cued locations. Experiment 2 demonstrated that when auditory stimuli consisted of a single tone, IOR did not modulate the facilitatory effect of sound, indicating that only cross-modal correspondence interacts with IOR and that alerting effects do not confound the results. Experiment 3 manipulated SOA to vary IOR magnitude and found that as IOR weakened, cross-modal correspondence at cued locations decreased and IOR' s modulatory effect diminished, supporting the applicability of the principle of inverse effectiveness to cross-modal correspondence.

Experiment 1' s interaction between IOR and cross-modal correspondence, according to additive factors logic (Sternberg, 1969), suggests that both factors operate at the same processing stage. Thus, our findings support the view that pitch-location cross-modal correspondence occurs at the perceptual processing stage, consistent with some previous research (Ković et al., 2010; Maeda et al., 2004). Some studies have argued for semantic-level processing (Gallace & Spence, 2006; Martino & Marks, 1999), proposing that shared semantic coding across modalities underlies cross-modal correspondence. However, the present findings demonstrate that cross-modal correspondence can occur at a purely perceptual level without requiring semantic mediation. This aligns with evidence that populations lacking "high-low" pitch terminology still exhibit pitch-location correspondence (Parkinson et al., 2012) and that pre-linguistic infants show cross-modal correspondence (Dolscheid et al., 2014; Walker et al., 2010). Our findings do not deny a potential role for semantic coding but suggest that for basic stimulus features like pitch and spatial location, which have natural correlations in the environment (e.g., heavier animals produce lower-frequency sounds and are less likely to be found at high elevations; Parkinson et al., 2012), perceptual correspondence can occur without semantic mediation. For stimulus features lacking natural correlations or for more complex stimuli, semantic coding may be necessary (Parise & Spence, 2012). Additionally, tasks encouraging semantic encoding may amplify semantic contributions to cross-modal correspondence (Martino & Marks, 1999).

Previous signal detection studies of loudness-brightness correspondence found that cross-modal correspondence affected decision criteria rather than perception (Marks et al., 2003). However, this paradigm differed from typical cross-modal correspondence tasks by requiring discrimination between two stimuli rather than detection of a single stimulus, potentially increasing reliance on decision processes. Different stimulus pairings may involve different mechanisms (Spence, 2011), and task demands may influence the locus of cross-modal correspondence effects.

Experiment 1' s finding that cross-modal correspondence occurred only at cued locations parallels some audiovisual integration studies. Previous research has found larger integration effects at cued locations during IOR (唐晓雨 et al., 2020), explained by the principle of inverse effectiveness, whereby weaker sensory input produces stronger integration (Meredith & Stein, 1983). This principle, initially discovered in single-neuron studies, has been confirmed in human behavioral and neural research (Rach et al., 2011; Senkowski et al., 2011; van de Rijt et al., 2019). It essentially reflects that weaker redundant information triggers stronger integration. Although cross-modal correspondence involves non-redundant stimuli, researchers have suggested that such stimuli may occupy the same end of a psychological continuum (e.g., high pitch and high location both represent the "high" end; McCormick et al., 2018), potentially sharing neural codes and producing redundancy. Thus, the principle of inverse effectiveness may apply to cross-modal correspondence. During IOR, reduced perceptual saliency at cued locations (Slagter et al., 2016) weakens visual input, producing stable cross-modal correspondence, whereas relatively higher saliency at uncued locations does not.

Experiment 3 directly confirmed the applicability of the principle of inverse effectiveness to cross-modal correspondence by showing that weakened IOR reduced cross-modal correspondence at cued locations and diminished IOR' s modulatory effect. This principle has been observed in both meaningless audiovisual integration (Senkowski et al., 2011) and multisensory speech perception (van de Rijt et al., 2019), where more difficult-to-perceive inputs produce greater multisensory enhancement. Our findings extend the principle' s applicability to cross-modal correspondence.

Cross-modal correspondence also influenced IOR. At 600-ms SOA, the IOR effect was smaller under congruent than incongruent conditions in both Experiments 1 and 3, as cross-modal correspondence increased target perceptual saliency, partially offsetting IOR-induced saliency reduction. At longer SOA, this difference disappeared because weakened cross-modal correspondence reduced its capacity to counteract IOR.

As the first investigation of IOR' s relationship with audiovisual cross-modal correspondence, this study needed to ensure that observed interactions truly reflected IOR' s modulation of cross-modal correspondence. Experiment 2, maintaining all other conditions from Experiment 1, manipulated sound presence to examine its pure effect. Because cross-modal correspondence requires relative mapping between two tones and two locations (Chiou & Rich, 2012), a single tone cannot produce correspondence. Results showed no interaction between IOR and sound presence, verifying that only cross-modal correspondence interacts with IOR and that alerting effects do not confound the findings. Although alerting effects can enhance perception (Kusnir et al., 2011) and interact with target saliency (Botta et al., 2017) and spatial attention (Botta et al., 2014; 2017), Experiment 2 found no interaction between alerting and exogenous-cue IOR. This may be because previous studies compared suprathreshold, threshold, and subthreshold stimuli (Botta et al., 2017; Chica et al., 2016), whereas our visual targets were fully visible at both cued and uncued locations, making IOR-induced saliency differences insufficient to trigger alerting modulation. Alternatively, alerting and IOR may operate via independent neural pathways. While both affect perceptual saliency (Botta et al., 2014; Prime & Jolicoeur, 2009) and involve frontoparietal networks (Bourgeois et al., 2012; Kusnir et al., 2011), alerting enhances perception top-down by amplifying input intensity through frontoparietal activation (Kusnir et al., 2011), whereas exogenous-cue IOR affects perception bottom-up by modulating input intensity and altering early visual projections to frontoparietal networks (Botta et al., 2014). Further neural investigations are needed to clarify the relationship between alerting and IOR. Nevertheless, the current study excludes alerting confounds, supporting the conclusion that pitch-location cross-modal correspondence occurs at the perceptual level.

6. Conclusion

IOR modulated audiovisual cross-modal correspondence, with stable cross-modal correspondence effects emerging at cued locations during IOR but not at uncued locations. Alerting effects induced by auditory stimuli did not interact with IOR. As IOR weakened, cross-modal correspondence at cued locations decreased, and IOR' s modulatory effect on cross-modal correspondence diminished. These results support the view that pitch-location cross-modal correspondence 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, 141–149.

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), 1256–1261.

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), 1456–1465.

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, 112–123.

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: Official Journal of the International Organization 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.

彭姓, 常若松, 李奇, 王爱君, 唐晓雨. (2019). 不同 SOA 下视觉返回抑制对视听觉整合的调节作用. 心理学报, 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: the official journal of the Society for 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., MacInnes, W. J., & Klein, R. M. (2021). Inhibition of return: An information processing theory of its natures and significance. *Cortex; a journal devoted to the study of the nervous system and behavior*, 135, 296-319.

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. (2020). 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), 373-373.

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), 1093-1103.

唐晓雨, 孙佳影, 彭姓. (2020). 双通道分配性注意对视听觉返回抑制的影响. 心理学报, 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.