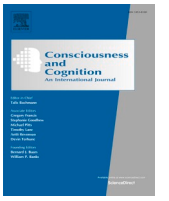
Consciousness and Cognition 122 (2024) 103709

Contents lists available at ScienceDirect

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/yccog



# Prior conscious experience modulates the impact of audiovisual temporal correspondence on unconscious visual processing

Hyun-Woong Kim [a,b], Minsun Park [c], Yune Sang Lee [a,d], Chai-Youn Kim [c,*]

[a] School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, United States
[b] Department of Psychology, The University of Texas at Dallas, Richardson, United States
[c] School of Psychology, Korea University, Seoul, Republic of Korea
[d] Department of Speech, Language, and Hearing, The University of Texas at Dallas, Richardson, United States

ARTICLE INFO

*Keywords:*
Continuous flash suppression
Consciousness
Visual awareness
Multisensory integration
Audiovisual synchrony
Causal inference

ABSTRACT

Conscious visual experiences are enriched by concurrent auditory information, implying audiovisual interactions. In the present study, we investigated how prior conscious experience of auditory and visual information influences the subsequent audiovisual temporal integration under the surface of awareness. We used continuous flash suppression (CFS) to render perceptually invisible a ball-shaped object constantly moving and bouncing inside a square frame window. To examine whether audiovisual temporal correspondence facilitates the ball stimulus to enter awareness, the visual motion was accompanied by click sounds temporally congruent or incongruent with the bounces of the ball. In Experiment 1, where no prior experience of the audiovisual events was given, we found no significant impact of audiovisual correspondence on visual detection time. However, when the temporally congruent or incongruent bounce-sound relations were consciously experienced prior to CFS in Experiment 2, congruent sounds yielded faster detection time compared to incongruent sounds during CFS. In addition, in Experiment 3, explicit processing of the incongruent bounce-sound relation prior to CFS slowed down detection time when the ball bounces became later congruent with sounds during CFS. These findings suggest that audiovisual temporal integration may take place outside of visual awareness though its potency is modulated by previous conscious experiences of the audiovisual events. The results are discussed in light of the framework of multisensory causal inference.

## 1. Introduction

Although our everyday visual environment is inundated with objects and events that give rise to an overflow of sensory signals, only a subset of the visual information can gain access to perceptual awareness. Concurrent auditory signals may influence perceptual selection of visual information by increasing perceived visual intensity (Chen et al., 2011; Stein et al., 1996), resolving perceptual uncertainty (Lewis & Noppeney, 2010; Sekuler et al., 1997), or strengthening a motion aftereffect (Park et al., 2019). Indeed, such audiovisual interaction promotes perceptual awareness of a visual representation congruent with auditory information when there is ambiguity in visual stimuli yielding bistable perception. For example, a concurrent sound biases visual perception of a bistable figure in favor of the interpretation that is congruent with the semantic information in the sound (Hsiao et al., 2012). In addition, during

* Corresponding author.
*E-mail address:* chaikim@korea.ac.kr (C.-Y. Kim).

https://doi.org/10.1016/j.concog.2024.103709
Received 17 January 2024; Received in revised form 9 May 2024; Accepted 14 May 2024
Available online 22 May 2024

binocular rivalry — perceptual alternations between dissimilar stimuli presented to the two eyes (Blake & Logothetis, 2002), a rival target becomes perceptually more dominant when it is accompanied by congruent sounds in terms of temporal synchrony (Conrad et al., 2013; Van Ee et al., 2009), motion direction (Conrad et al., 2010), and musical melodies (Kim et al., 2017; Lee et al., 2015).

The effectiveness of audiovisual interaction in enhancing dominance of perceptually selected visual stimuli naturally invites a question of whether congruent auditory information can also influence visual processing for those that are not selected for awareness. That is, can audiovisual interaction take place between an audible sound and a visual stimulus presented outside of awareness such that the sound boosts conscious access to the *invisible* visual stimulus? This question may be addressed using continuous flash suppression (CFS), a variant of binocular rivalry wherein a flashing visual mask with a high contrast is presented to one eye to produce prolonged perceptual suppression of a more steady, lower-contrast visual stimulus presented to the other eye (Tsuchiya & Koch, 2005). Typically, a visual stimulus is rendered invisible (i.e., outside conscious awareness of an observer) by a CFS mask while an audible sound — congruent or incongruent with the visual stimulus — is concurrently presented. The latency until the visual stimulus breaks through CFS is compared between audiovisual congruent and incongruent conditions, with a faster detection in a congruent than in an incongruent condition taken as evidence of audiovisual interactions in the absence of visual awareness (Yang et al., 2014). For instance, spatial correspondence between audible sounds and invisible flash stimuli has been shown to boost visual detection (Aller et al., 2015; Delong et al., 2018). A recent study also found that a motion aftereffect induced by adaptation to invisible gratings moving leftward or rightward was strengthened when accompanied by a congruent direction of auditory motion during adaptation (Park et al., 2024). In addition, for audiovisual natural scenes, the time for breaking suppression was shorter when a scenery soundtrack was semantically congruent with a visual scene under CFS, compared to when it was incongruent (Cox & Hong, 2015; Tan & Yeh, 2015).

In contrast to the evidence of audiovisual interaction during visual suppression based on spatial or spatial-semantic properties, there are mixed results in studies using temporal correspondence between auditory and visual stimuli. Hong and Shim (2016) found using temporal synchrony between flickers and beeps that the audible beep sounds temporally synchronous to the flickering visual target under CFS modulated the detection time of the visual target. However, Moors et al. (2015) reported that temporal congruency between a visual looming stimulus and a beep/looming sound had no impact on the contrast detection threshold of the visual stimulus under CFS. From a classical point of view, such an inconsistency is counterintuitive in that temporal coincidence is a crucial principle for multisensory perception to occur (Vroomen & Keetels, 2010). Moreover, it is challenging to be reconciled with the extant findings that audiovisual interaction can transpire outside of visual awareness based on apparently more complex temporal synchrony between

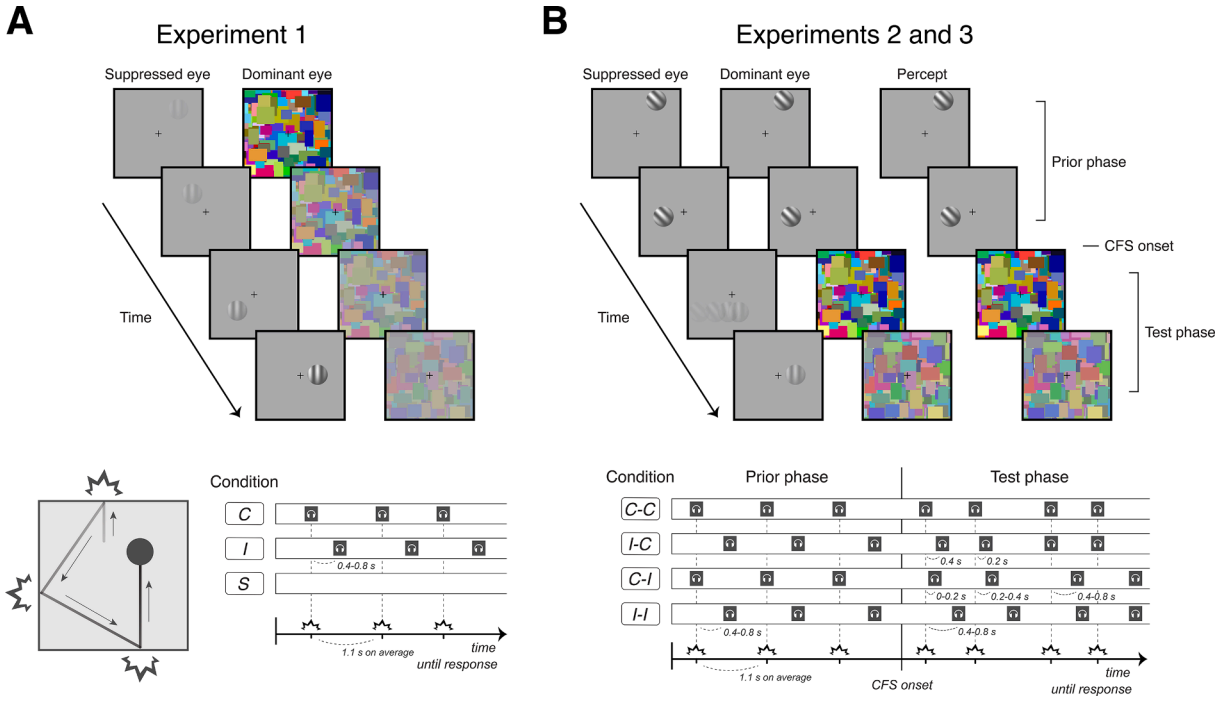



**Fig. 1.** Stimuli and procedures. (top) Schematics of stimulus sequence in a single trial. (A) In Experiment 1, a CFS mask was presented along with a ball-shaped object at the onset of each trial. (B) In Experiments 2 and 3, there was an initial prior phase where a ball-shaped object was presented to both eyes before a CFS mask was presented to a dominant eye during a subsequent test phase. The grating pattern on the ball surface was gradually rotated from a diagonal to horizontal or vertical direction following the CFS onset. (bottom) Schematics of stimulus conditions. Spiky symbols indicate the timing of visual collision while headphone symbols indicate sound timings. (A) In Experiment 1, the visual display was accompanied by click sounds that were temporally matched (congruent, *C*) or unmatched (incongruent, *I*) to bounces of the ball, or slience (*S*). (B) In Experiments 2 and 3, the temporal synchrony between bounces and sounds was independently manipulated preceding (prior phase) and following (test phase) the onset of the CFS mask, comprising four conditions: *C-C*, *I-C*, *C-I*, and *I-I*.

speech sounds and lip movements (Alsius & Munhall, 2013; Plass et al., 2014). In these studies, a speech soundtrack congruent with dynamic lip movements suppressed by a CFS mask expedited the time to detect the suppressed visual stimulus (Alsius & Munhall, 2013) or the time to discriminate spoken words (Plass et al., 2014).

The mixture of evidence may be because not all proximal multimodal signals result from the same origin in reality; distinct events may produce their respective sensory signals simultaneously with no apparent relationship. Indeed, the brain does not just passively bind multimodal percepts that are temporally proximal but makes active inferences about the causal structure to correctly integrate those sharing the same source and to segregate those occurring together by coincidence (Körding et al., 2007; Noppeney, 2021). Moreover, perceived causality between visual and auditory events has been shown to dissociate from their perceived simultaneity, with the former influencing the latter by shifting or widening audiovisual temporal binding window (Kohlrausch et al., 2013; Vroomen & Keetels, 2020). Given the role of causal inference in audiovisual temporal integration, prior inference about the causal structure governing the individual instances of temporal correlation between visual and auditory events may be necessary for those instances to be integrated outside of visual awareness. From this standpoint, it is important to note that the temporal congruency of audiovisual looming cues used in Moors et al. (2015) is based on the causal structure of looming phenomenon that gives rise to synchronous changes in both visual and auditory signals, i.e., getting larger in apparent size and louder in amplitude respectively. As such, those signals might have failed to be integrated outside of visual awareness since the audible stimulus (i.e., a tone gradually increased in loudness) was not specific enough to inform its causal relevance to the invisible visual stimulus (i.e., a concentric grating gradually increased in apparent size). Such a causal inference account for audiovisual interactions without visual awareness is not incompatible with other existing evidence, in that the audiovisual congruency in Hong and Shim (2016) — i.e., synchronous appearances of flickers and beeps — does not rely on a specific causal structure (e.g., looming) that would need to be inferred for subsequent integration whereas the audible speech soundtrack in Alsius and Munhall (2013) and Plass et al. (2014) is causally specific to the lip movements presented outside of awareness (see General discussion for further explanation).

In the present study, we investigated conditions under which a concurrent audible sound having a causal relation to a visual stimulus can be integrated with the visual stimulus presented outside of awareness using CFS. An audiovisual event involving a causal structure that elicits a temporally synchronous percept is an object colliding with a hard surface, producing a brisk contact sound. Here we used an animation of a ball constantly moving inside a square frame window and bouncing off four boundaries of the frame, which was perceptually suppressed using a CFS mask (Fig. 1A). In Experiment 1, every moment that the ball contacted one of the square boundaries was accompanied by one of the following: a temporally congruent click sound (congruent: *C*), an incongruent sound with a random delay (incongruent: *I*), or no sound (silent: *S*) (Fig. 1A). We measured the latency to detect a visually suppressed ball stimulus in each stimulus condition to assess the impact of audiovisual temporal integration on the processing of the invisible ball stimulus, which was estimated by the difference in mean detection time between congruent and incongruent/silent conditions. To foreshadow the results, we failed to find evidence of audiovisual integration outside of visual awareness: detection of the ball breaking interocular suppression did not differ when accompanied by congruent sounds from incongruent or no sounds.

Setting out from this null finding, Experiments 2 and 3 were designed to elucidate the role of prior conscious audiovisual experience on the subsequent audiovisual temporal integration without visual awareness. Specifically, we examined whether a preview of temporal congruency between a moving ball and contact sounds modulates audiovisual integration between them when the ball stimulus was later rendered invisible by CFS (Fig. 1B and Movie 1 in Appendix A). For this purpose, we manipulated the audiovisual temporal congruency (i.e., congruent or incongruent) both before and while the ball stimulus is suppressed by a CFS mask, resulting in four audiovisual stimulus conditions: congruent-congruent (*C–C*), incongruent-congruent (*I-C*), congruent-incongruent (*C-I*), and incongruent-incongruent (*I-I*) (Movie 2). For instance, in the *I-C* condition, a 'visible' period of temporally incongruent audiovisual events (*I*) was followed by an 'invisible' period of the same events but with temporal congruency (*C*) (see Methods for more details).

This manipulation allowed us to examine not only (1) whether prior experience of temporally congruent audiovisual stimuli promotes audiovisual integration between them in the absence of visual awareness, but also (2) whether such an unconscious integration is disrupted when the audiovisual causality is not established due to prior experience of temporally incongruent audiovisual stimuli. Accordingly, we formed two specific hypotheses: (1) congruent sounds accompanying the ball stimulus for both before and after the onset of CFS (i.e., *C–C*) would facilitate the detection of the ball breaking suppression compared to when incongruent sounds are presented during CFS (i.e., *C-I* and *I-I*), and (2) prior experience of temporally incongruent audiovisual stimuli would disrupt visual detection of the ball even when congruent sounds are presented during CFS (i.e., *I-C*), which would be manifested as a longer mean detection time in the *I-C* compared to *C–C* conditions. The prior audiovisual congruency information was passively given in Experiment 2 while it was explicitly monitored in Experiment 3, which allowed us to estimate the role of both implicit and explicit knowledge of audiovisual temporal congruency in the degree of audiovisual integration without visual awareness. The results from the three experiments are reported below.

## 2. Experiment 1

### 2.1. Methods

#### 2.1.1. Participants

Twenty-four individuals participated in Experiment 1 (mean age: 23.7 years, 17 females). All participants reported normal or corrected-to-normal visual acuity and normal stereopsis. They gave written informed consent approved by the Institutional Review Board of Korea University (1040548-KU-IRB-17–85-A-1).

*2.1.2. Stimuli and procedures*

All stimuli were generated using MATLAB equipped with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Visual stimuli were presented on a linearized 17-inch CRT monitor (1024 x 768 resolution, 60-Hz refresh rate, 40 cd/$m^2$ mean luminance) and were viewed dichoptically through a mirror-stereoscope with the head stabilized by a head/chin rest. Auditory stimuli were delivered through SRH440 headphones. All experiments were conducted in a quiet, dark room.

For a CFS mask, we used 255 high-contrast and colored Mondrian patterns, each of which comprised rectangles of random size (0.4° – 0.6° visual degree in length) and random color (Fig. 1). These patterns were normalized to 75 % RMS contrast and presented at 10 Hz. The CFS mask was presented within a square window (3° x 3°) surrounded by black lines (0.1° in thickness). The target stimulus was a sinusoidal grating (spatial frequency of 2.7 cycles per visual degree) centered within a circular window (0.75° in diameter). The brightness of the grating was filtered through a cosine mask centered at upper-left position of the circular window to make shade on the grating such that it looked like a 'ball object' (Fig. 1). The ball stimulus was presented on a gray background confined by the same square frame as the one for the CFS mask. It moved in a straight line with a constant velocity of 1.5° per second and bounced off the inner boundaries of the frame window. The two square frames, each of which displayed the CFS mask and the ball stimulus, were presented dichoptically to two eyes. A bubble-shaped figure was displayed on the background of each frame window to support stable binocular alignment. A small, black cross at the center of each window provided a fixation point during the experiment. For an auditory stimulus we used a click sound of 10-ms duration and 440 Hz in tonal frequency.

Each trial began when a participant pressed the space bar. A CFS mask was immediately presented to the dominant eye, while at the same time a ball target moving on a gray background was presented to the non-dominant eye. The dynamics of the ball stimulus were characterized by horizontal/vertical and diagonal movements. The ball stimulus moved straightly in one of four directions (up, down, right, and left) at the beginning and, after the first bounce, headed in either clockwise or counterclockwise direction with straight diagonal motion (see Movie 1 in Appendix A). Each bounce position of the ball was randomly determined within the middle half of the four boundaries within the square frame, meaning that the motion path of the ball was different from trial to trial. The interval between two adjacent bouncing moments was 1.11 s on average (range: 0.57 to 1.62 s).

The grating on the ball surface was oriented either horizontally or vertically. The grating contrast was set to a value between 5 % and 10 %; this initial contrast value was determined for each of the individual participants based on a staircase procedure during practice before the main experiment (see below). The initial contrasts of both the ball stimulus and the CFS mask remained the same for 3 s, and then were gradually ramped up (20 %/s in log unit) and down (10 %/s in log unit) respectively, lasting up to 9 s. Participants were instructed to press a down-arrow key as soon as they detected any motion of the ball breaking through the flashing colorful patterns. They were also asked to hold the button down while they were unsure about the orientation of the grating, and then to release the button when it became discriminable. Each test phase lasted until the release of detection button, or for a maximum duration of 12 s. Upon releasing the button participants were prompted to report whether the orientation of the grating was horizontal or vertical by pressing left or right arrow key.

There were three stimulus conditions — two audiovisual and one silent. In the two audiovisual conditions, the visual display was accompanied by click sounds, which were temporally matched (congruent) or unmatched (incongruent) to the timing of the ball bouncing off the inner square boundaries (Fig. 1A). In the incongruent condition, the click sounds were delayed randomly for between 400 and 800 ms with respect to the bounces. The auditory stimuli were presented until participants made a detection response. In the silent condition, there were no click sounds presented throughout a task trial. Participants were told that, most of times, they would hear some sounds during the experiment but were provided no information about the audiovisual congruency.

In addition, there were 10 catch trials for each condition. During a catch trial, only a gray background was initially presented to the non-dominant eye, and after a random duration of time (determined based on participants' response history during previous task trials), a ball target was superimposed over the CFS mask presented to the dominant eye with a moderate opacity as though it broke through suppression. The catch trials were included to verify that participants faithfully responded following visual detection and to

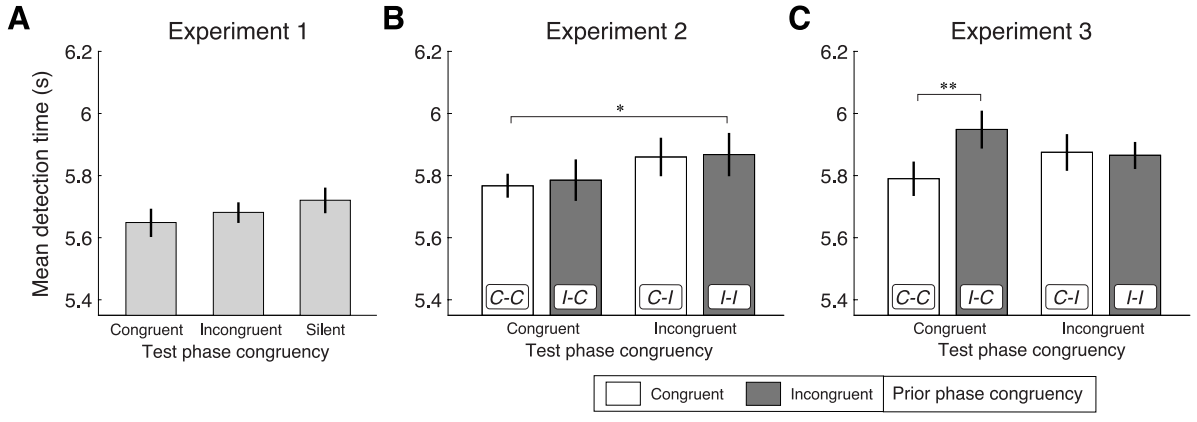



**Fig. 2.** Mean detection time in Experiments 1, 2, and 3 as a function of prior and test phase audiovisual congruency. Asterisks indicate significant differences in pairwise comparisons. Error bars denote 95% within-subject confidence intervals (Loftus and Masson, 1994).

rule out the possibility that any audiovisual congruency effect is due to response bias. Each stimulus condition was repeated 60 times, resulting in a total of 210 trials including 30 catch trials presented in a randomized order.

Before the main experiment, participants underwent a staircase procedure to determine individual target contrast levels, which also served as practice for the main portion. Participants completed 18 task trials during this procedure. The ball contrast was sequentially reduced with a 1 % step from 10 % to a minimum of 5 % to lengthen the duration of suppression, when the ball detection time (i.e., suppression time) was shorter than 5 s. The contrast was sequentially increased when the target was not detected throughout the CFS phase, up to 15 %. The individual contrast levels ranged from 5-10 %, with a mean of 8.7 %. There was no sound during the practice.

For each participant, incorrect trials and no-breakthrough trials were excluded from the analysis (1.2 % on average). We also excluded individual trials with the ball detection times deviated ± 3 SD from the mean detection time for each stimulus condition (0.05 % on average).

### 2.2. Results and discussion

Fig. 2A shows mean detection times of the ball stimulus from CFS for the three stimulus conditions in Experiment 1. A repeated measures ANOVA indicated that the effect of the stimulus condition was not statistically significant [$F(2, 46) = 1.72$, $p = 0.190$, $\eta_p^2 = 0.070$], although the mean detection time was slightly shorter when accompanied by audiovisual congruent sounds compared to incongruent sounds or silence. Given the non-significant result, we additionally performed a Bayes factor (BF) analysis using ttestBF function of the *BayesFactor* package (Rouder et al., 2009), to estimate the degree of evidence towards or against the null hypothesis. The BF for the comparison between audiovisual congruent and incongruent conditions was 0.22, indicating moderate evidence in favor of the null model according to Lee and Wagenmakers (2014). The BF for audiovisual congruent vs. silent conditions was 0.61, indicative of anecdotal evidence towards the null model.

During the catch trials, participants reported visual detection after target presentation for 95.8 % of the trials on average, indicating that they faithfully performed the detection task. To assess the latency of button response upon visual detection, reaction times from the onset of a catch stimulus (i.e., a ball stimulus superimposed on the CFS mask) were calculated and averaged for each stimulus condition. The mean reaction time was not different across the congruent ($M = 620$ ms), incongruent ($M = 624$ ms), and silent conditions ($M = 629$ ms) [$F(2, 46) = 0.14$, $p = 0.872$, $\eta_p^2 = 0.006$], indicating that participants' responsiveness to the visual target was not affected by concurrent congruent or incongruent sounds.

The results from Experiment 1 suggest a lack of temporal integration between congruent audiovisual signals when the visual stimulus is suppressed from awareness. However, as discussed in Introduction, this null finding might be because participants may have failed to make inferences about the causal structure (i.e., collision causing a contact sound) underlying the specific instances of visual and auditory events in the present experiment (i.e., synchronous ball bounces and click sounds), making the integration less likely to occur in the absence of visual awareness. To address this possibility, in Experiment 2, we examined whether audiovisual temporal integration outside of visual awareness could transpire from prior conscious experience of temporally congruent audiovisual events, which would inform their causal relevance and thereby allow causal inferences.

## 3. Experiment 2

### 3.1. Methods

#### 3.1.1. Participants

Twenty-five individuals participated in Experiment 2 (mean age: 22.7 years, 11 females). All participants reported normal or corrected-to-normal visual acuity and normal stereopsis. They gave written informed consent approved by the Institutional Review Board of Korea University (1040548-KU-IRB-17–85-A-1).

#### 3.1.2. Stimuli and procedures

In Experiment 2, we used the same visual and auditory stimuli as those used in Experiment 1 but included an additional phase before the CFS mask was presented. During this prior phase, a ball target moving inside the square frame window was presented to both eyes. The grating on the ball surface was oriented + 45° or –45° from vertical, and its contrast was set moderately high (25 %-35 %) to ensure its visibility. This phase lasted 3–4 s, during which the ball stimulus bounced off the frame boundaries three times with clockwise or counterclockwise diagonal movements (see Movie 1). The next, test phase began with the onset of a CFS mask presented to the dominant eye just before the fourth bounce, which made the ball target being presented to the non-dominant eye immediately disappear from visual awareness (Fig. 1B). The interval duration between the CFS onset and the fourth bounce was randomly chosen between 200 and 400 ms. The ball made a horizontal/vertical movement after the third bounce, and then made diagonal movements from the fourth bounce until the end of the test phase, the direction of which was either clockwise or counterclockwise again. This sequence of diagonal-horizontal/vertical-diagonal movements around the CFS onset was designed to make unpredictable the direction of its motion (i.e., clockwise or counterclockwise) during the test phase. The surface grating was gradually rotated into a horizontal or vertical orientation in the next 500 ms following the CFS onset. To assure a complete suppression during the initial period of the test phase, the visual contrast decreased to 0 % at the CFS onset and recovered in next 500 ms to a value between 5 % and 15 %, which was determined during the practice (range: 9–15 %, with a mean of 10.4 %). The rate of incorrect/no-breakthrough trials was 2.1 % on average, and the rate of trials with the ball detection times deviated ± 3 SD from the mean detection time for each stimulus condition was 0.06 % on average. These trials were excluded from data analyses.

The temporal congruency between bouncing movements and contact sounds was independently manipulated preceding (prior congruency) and following (test congruency) the onset of CFS, resulting in four different audiovisual stimulus conditions (Fig. 1B and Movie 2 in Appendix A). The bounce-sound timing was matched across the two phases in the 'congruent-congruent' (*C–C*) condition whereas the sounds were delayed randomly for between 400 and 800 ms with respect to the bounces in the 'incongruent-incongruent' (*I-I*) condition. In the 'incongruent-congruent' (*I-C*) and 'congruent-incongruent' (*C-I*) conditions, the audiovisual congruency was changed following the CFS onset by a gradual decrease (*I-C*) or increase (*C-I*) of the delay interval over the next two bounces respectively (see Fig. 1B). This transition period was included to mitigate abrupt changes in time intervals between adjacent sounds around the CFS onset. We asked participants to observe the ball movement during the prior phase, again without providing any information about the audiovisual congruency. Each stimulus condition was repeated 60 times without catch trials, resulting in a total of 240 trials. The practice session was reduced to 12 trials. Otherwise, the experimental procedures were the same as those used in Experiment 1.

### 3.2. Results and discussion

Fig. 2B shows mean detection times of the ball stimulus from CFS for the four stimulus conditions. A two-way repeated measures ANOVA of mean detection time with the factors of prior and test phase congruency revealed a main effect of test congruency [$F(1, 24) = 9.69$, $p = 0.005$, $\eta_p^2 = 0.288$], indicating that bouncing-ball-under-CFS was detected faster when accompanied by congruent sounds (i.e., *C–C* and *I-C*) compared to incongruent sounds (i.e., *C-I* and *I-I*). In contrast, neither the main effect of prior congruency [$F(1, 24) = 0.18$, $p = 0.676$, $\eta_p^2 = 0.007$] nor the interaction [$F(1, 24) = 0.01$, $p = 0.910$, $\eta_p^2 = 0.001$] was significant, indicating that the detection time was not affected by the audiovisual congruency between ball bounces and contact sounds during the prior phase.

To further examine our hypotheses, we computed BFs for the planned contrasts between the *C–C* condition and other incongruent conditions during the test phase (i.e., *C-I* and *I-I*) as well as between the *C–C* and the other congruent condition where audiovisual stimuli were incongruent during the prior phase (i.e., *I-C*). The former contrasts (*C–C* vs. *C-I* and *I-I*) indicated moderate-to-strong evidence for an effect of audiovisual congruency during the test phase on the mean detection time (BF = 9.89), whereas the latter contrast (*C–C* vs. *I-C)* indicated moderate evidence for a null effect (BF = 0.16).

In Experiment 2, we found a facilitating effect of temporally congruent sounds on the processing of a suppressed visual stimulus, indicating that the temporal congruency between bounces and sounds boosted the ball into visual awareness. This result is deviated from that of Experiment 1 where there was no detection advantage of audiovisual congruency without a prior phase. This suggests that audiovisual temporal integration can take place in the absence of visual awareness given prior conscious experience of the audiovisual events, supporting our first hypothesis. By contrast, against our second hypothesis, the mean detection time was comparable between the *C–C* and *I-C* conditions, suggesting that prior experience of temporally incongruent audiovisual events may not have disrupted the subsequent integration between temporally congruent audiovisual events during CFS.

Nevertheless, there are two considerations to note. First, it is uncertain whether prior audiovisual congruency was explicitly processed during the prior phase because the concurrent sounds were completely irrelevant to the visual detection task. Second, despite the random delay between every bouncing event and sound in the incongruent condition, there was some degree of temporal proximity between the audiovisual events (see Fig. 1B). Thus, both congruent and incongruent audiovisual stimuli might have been regarded as being causally related at an implicit level. Indeed, none of the participants in Experiment 2 reported awareness of audiovisual temporal congruency of individual trials when asked informally after participation. In Experiment 3, we examined whether explicit processing of prior audiovisual congruency modulates later integration without visual awareness.

## 4. Experiment 3

### 4.1. Methods

#### 4.1.1. Participants

Thirty individuals participated in Experiment 3 (mean age: 21.8 years, 16 females). All participants reported normal or corrected-to-normal visual acuity and normal stereopsis. They gave written informed consent approved by the Institutional Review Board of Korea University (1040548-KU-IRB-17–85-A-1).

#### 4.1.2. Stimuli and procedures

The overall procedure was identical to that of Experiment 2 except the addition of a retrospective audiovisual congruency judgment task on the stimuli during the prior phase right after the ball detection task. Specifically, participants were asked to discriminate whether the sounds accompanied by the visual display were temporally matched to bounces of the ball during each prior phase, and to report it by pressing the left (unmatched) or right (matched) arrow button after they made a detection response in the following test phase. This dual task procedure was intended to require explicit processing of temporal relationship between the audiovisual stimuli. Each stimulus condition was repeated 48 times, for a total of 192 trials. There were two phases of practice sessions. The first was the same as the one in Experiment 2. In the second phase, participants listened to congruent or incongruent sounds to practice audiovisual congruency judgments during the prior phase. Each of the two practice phases consisted of 8 trials. Otherwise, the experimental procedures were the same as those used in Experiment 2.

The individual contrast levels determined during the practice ranged from 6-10 %, with a mean of 9.0 %. The rate of incorrect/no-breakthrough trials was 2.7 % on average, and the rate of trials with the ball detection times deviated ± 3 SD from the mean detection

time for each stimulus condition was 0.05 % on average. These trials were excluded from data analyses.

### 4.2. Results and discussion

Although prior audiovisual congruency was readily discernible to participants, the discrimination performance was not perfect (mean: 90 %), probably because of the failure to recall a correct answer after making a detection response. The performance was not significantly different between conditions ($p > 0.113$).

Fig. 2C shows mean detection times of the ball stimulus from CFS for the four stimulus conditions. A repeated measures ANOVA showed a significant main effect of prior congruency [$F(1, 29) = 5.33$, $p = 0.028$, $\eta_p^2 = 0.155$]. By contrast, the main effect of test congruency was not significant [$F(1, 29) < 0.01$, $p = 0.980$, $\eta_p^2 < 0.001$]. Critically, we found a significant interaction effect between the two factors [$F(1, 29) = 8.87$, $p = 0.006$, $\eta_p^2 = 0.234$]. Pairwise comparisons revealed that the effect of prior congruency was shown only when audiovisual stimuli were congruent during the test phase [$t(29) = 3.44$, $p = 0.002$], with a longer mean detection time for the *I-C* compared to *C–C* conditions (Fig. 2C).

In line with ANOVA, a BF analysis on the *C–C* vs. *I-C* contrast yielded strong evidence for the difference between the two conditions (BF = 17.78). This finding supports our second hypothesis, suggesting that explicit processing of *temporally incongruent* audiovisual stimuli during the prior phase hampers the integration between *temporally congruent* audiovisual signals during the later test phase in the absence of visual awareness. However, unlike Experiment 2, the first hypothesis was not supported by the results from Experiment

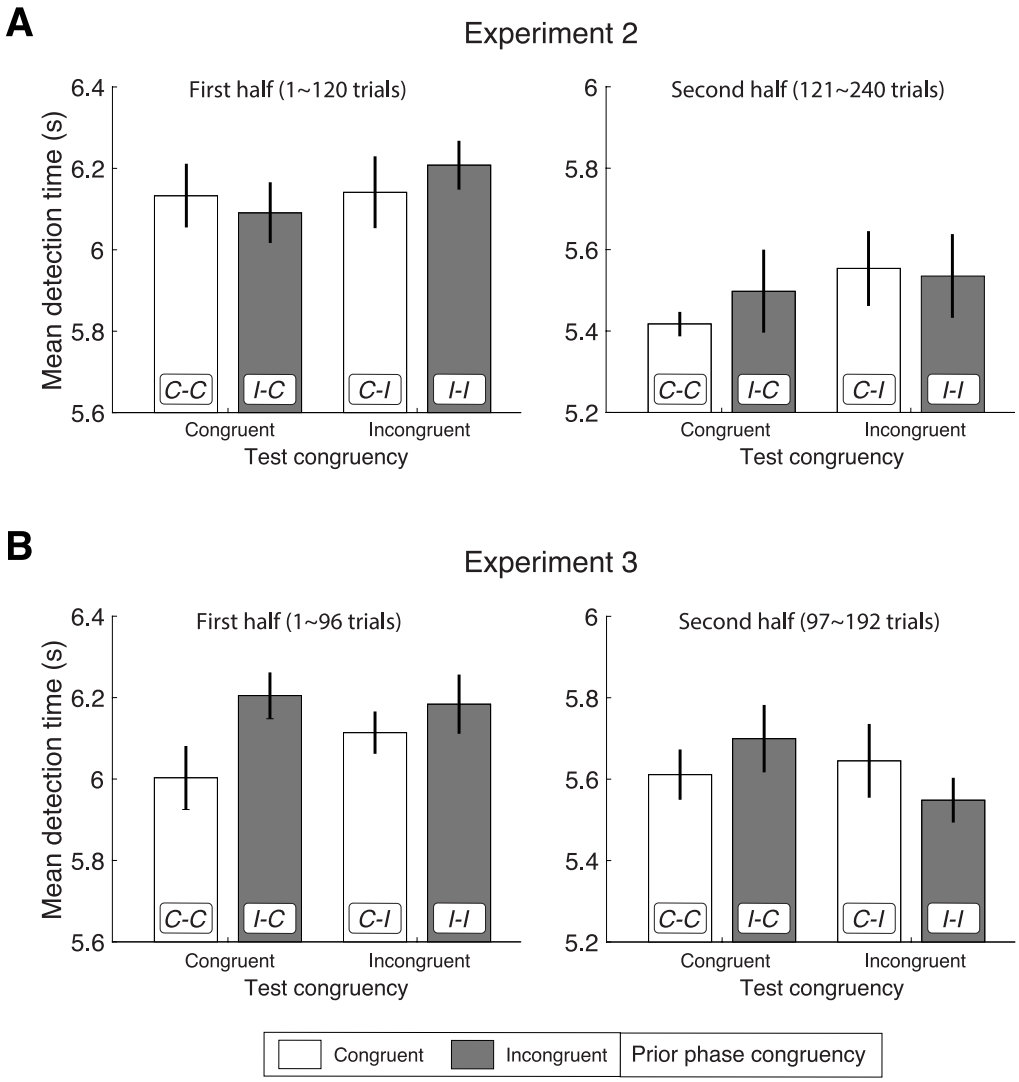



**Fig. 3.** Mean detection time from the median-split data in Experiments 2 and 3 as a function of prior and test phase audiovisual congruency. Error bars denote 95% within-subject confidence intervals.

3: although the mean detection time tended to be shorter in the *C–C* compared to *C/I-I* conditions (Fig. 2C), the BF on the *C–C* vs. *C/I-I* contrast indicated lack of significant evidence for an audiovisual congruency effect during the test phase (BF = 0.71).

One possible reason for the lack of test audiovisual congruency effect in Experiment 3 may come from the fact that not only congruent but incongruent audiovisual events were also explicitly processed during the prior phase. According to causal inference models, the perceptual system cumulatively computes and updates the probabilities of causal structures: whether audiovisual events are causally related or not (Kayser & Shams, 2015; Shams & Beierholm, 2010). In the context of our study, the probability of the causal relationship between audiovisual events during the prior phase was 50 %, in that they were incongruent in half of the trials. Thus, the explicit confirmation of audiovisual congruency in Experiment 3 may have generally weakened the strength of the causal connection of concurrent sounds to ball bounces as a consequence of visual collisions. Such an active inference in favor of segregation may have made audiovisual temporal integration less likely to occur without visual awareness, leading to a less reliable test congruency effect under CFS.

Would cumulative explicit experiences of incongruent audiovisual events weaken their interaction outside of visual awareness? If so, the size of the test audiovisual congruency effect should be reduced over the course of the experimental session. In addition, the prior audiovisual congruency effect is likely to be reduced as well due to the weakening of audiovisual integration in the *C–C* condition over time. To examine this possibility, we divided the task trials from Experiment 3 into two halves and performed separate BF analyses. Indeed, the BF on the test congruency effect (i.e., the *C–C* vs. *C/I-I* contrast) was decreased from the first (BF = 2.67) to the second half (BF = 0.15) as shown in Fig. 3B. Similarly, the BF on the prior congruency effect (i.e., *C–C* vs. *I-C* contrast) was also markedly reduced from the first (BF = 32.12) to the second half (BF = 0.42). This is in stark contrast with the same analysis from Experiment 2, where the BF on the test congruency effect was substantially increased from the first (BF = 0.21) to the second half (BF = 11.42, see Fig. 3A). The inconsistency in the results between the three experiments is discussed further in the general discussion.

## 5. General discussion

The present study was set out to investigate in which conditions cross-modal interaction between temporally congruent auditory and visual stimuli can transpire outside of visual awareness. The experiments reported here show that an unconscious temporal integration of audiovisual stimuli relies strongly on an observer's prior conscious experience of those. In Experiment 1, where no experience was given prior to interocular suppression using a CFS mask, we found little evidence of audiovisual temporal integration without visual awareness. In contrast, in Experiment 2, a brief pre-exposure to auditory and visual events before undergoing CFS yielded a significant impact of audiovisual temporal congruency on unconscious visual processing by promoting access to perceptual awareness. This finding suggests that audiovisual stimuli, at least for those used here, need to be consciously experienced first, in order for them to be temporally integrated later at an unconscious level. In Experiment 3, we further demonstrated that explicit processing of temporally incongruent audiovisual stimuli before undergoing CFS resulted in a significant slowing of visual detection when those stimuli were congruent later during CFS. However, the prior explicit processing of audiovisual temporal relationship substantially weakened the effectiveness of audiovisual congruency under CFS in expediting visual access. In the following we discuss potential implications of the significant as well as non-significant results obtained from this study.

As outlined in Introduction, previous research has demonstrated inconsistent evidence regarding whether audiovisual interaction based on temporal congruency can take place under CFS. While concurrent speech sound has been shown to expedite perceptual awareness of suppressed lip movements (Alsius & Munhall, 2013; Plass et al., 2014), a seemingly simpler temporal correspondence between audiovisual looming signals did not affect processing of a suppressed visual looming stimulus (Moors et al., 2015). The current study provides a clue to understanding this discrepancy: unlike speech sounds that have a specific causal relevance to lip movements, relatively simpler and less specific auditory stimuli such as looming or click sounds in an experimental setting may require prior conscious experience of intact audiovisual events to enable causal inference about whether or not those sounds causally relate to visual stimuli, so as to be later integrated in the absence of visual awareness. Consistent with this idea, Faivre et al. (2014) have reported that subliminal priming of semantic congruency between audiovisual word pairs impacted supraliminal processing of congruency of target word pairs, but only when the priming stimuli had been consciously processed during a previous training session. It is an open question the extent to which suprathreshold auditory signals have to be causally specific to visual signals to result in audiovisual interaction outside of visual awareness with no prior conscious experience.

A similar line of reasoning applies to explain the results from Hong and Shim (2016), where a prior experience to audiovisual stimuli was not necessary to obtain a reliable audiovisual congruency effect on visual processing under CFS using a flickering visual stimulus accompanied by beep sounds in synchrony with the flickering. As mentioned in Introduction, this form of audiovisual temporal correspondence — simultaneous pop-ups of auditory and visual stimuli — does not rely on a specific causal structure underlying the temporal correlation of concurrent audiovisual events like looming or collision. Moreover, such kind of audiovisual temporal integration could possibly take place at earlier stages of visual hierarchy such as superior colliculus (Holmes & Spence, 2005; Meredith & Stein, 1983). Relatedly, auditory beeps synchronous with instantaneous color changes of visual objects have been shown to automatically draw attention to the visual changes, improving visual search performance (Van Der Burg et al., 2008). Given the pre-attentive nature, audiovisual events involving simultaneous presentations may not require prior experience for a cross-modal interaction to occur without visual awareness.

Although pre-exposure to audiovisual stimuli led to a faster detection of the later suppressed visual stimulus when accompanied by congruent compared to incongruent sounds in Experiment 2, the audiovisual congruency effect during the test phase was not modulated by whether the audiovisual stimuli presented during the prior phase were congruent or incongruent. However, as discussed earlier, the incongruent sound sequences had some degree of temporal correlations with bouncing moments; this was intentional to

ensure the ambiguity of audiovisual congruency during the test phase as well as to enable smooth transitions from congruent to incongruent conditions and vice versa. Due to the temporal correlation, the slight asynchrony between the ball movements and the contact sounds during the prior phase could have broadly informed their causal relevance at an implicit level rather than their precise temporal relation on a trial-to-trial basis, based on the causal structure of collision-sound events that can be inferred from everyday perceptual experiences. Indeed, we found a significant *prior* audiovisual congruency effect when the audiovisual temporal relationship during the prior phase was explicitly processed in Experiment 3, with a slower visual detection for temporally congruent audiovisual stimuli after explicit processing of incongruent ones. This result is consistent with the previous literature suggesting the importance of attention and task relevance in multisensory perception (Talsma et al., 2010) and causal inference (Noppeney, 2021). For instance, a recent study showed that explicit self-report on spatial representations of audiovisual percepts influenced whether the audiovisual signals would be integrated or segregated (Ferrari & Noppeney, 2021).

One may argue that the prior audiovisual congruency effect observed in Experiment 3 could be due to a processing cost associated with expectation violation. In line with this idea, there is evidence that a suppressed visual stimulus is detected earlier when it was expected than when unexpected by a pre-stimulus cue valid for 74 % and 10 % of task trials in the expected and unexpected conditions, respectively (Pinto et al., 2015). However, this account is limited in that there was no difference in detection time for incongruent audiovisual stimuli under CFS between after incongruent ones (i.e., *I-I*) and after congruent ones (i.e., *C-I*), the latter of which should also involve an expectation violation. The current finding may be better explained within the framework of multisensory causal inference, which posits that previous causal inferences determine whether to integrate or segregate multimodal signals in spatio-temporal proximity (Kayser & Shams, 2015; Körding et al., 2007; Noppeney, 2021). Previous research has shown that perceptual segregation between audiovisual speech signals is promoted by an incongruent audiovisual context where lip movements were paired with an incongruent syllable sound, compared to a congruent audiovisual speech context (Gau & Noppeney, 2016). Similarly, in Experiment 3, the explicit processing of incongruent audiovisual stimuli in the prior incongruent condition may have resulted in segregation between the auditory and visual inputs, disrupting the integration of congruent audiovisual events without visual awareness during the later test phase. From this view, our finding suggests that visual information under the surface of awareness can still interact with auditory inputs as a function of prior audiovisual experiences, granting it greater or lesser potential to be selected for perceptual awareness.

Although we observed a numerically faster mean detection time for congruent (i.e., *C–C*) compared to incongruent (i.e., *C-I* and *I-I*) audiovisual stimulus conditions during the test phase in Experiment 3, the difference was not statistically significant, unlike in Experiment 2. This result suggests that audiovisual causal relation has been weakened due to cumulative explicit experiences of temporally incongruent audiovisual events. Supporting this idea, both prior and test audiovisual congruency effects in Experiment 3 diminished from the first to the second half of the trials. Intriguingly, the opposite pattern emerged in Experiment 2, where a stronger effect of test audiovisual congruency was obtained in the second than the first half of the trials. We speculate that this finding reflects a cumulative effect of implicit causal inferences on the audiovisual events during the prior phase. Together, these results show that the audiovisual congruency effects on visual detection dynamically evolve as prior experience accumulates, adding further support to a causal inference account.

One limitation of the current study is that since there was no baseline condition in Experiments 2 and 3, we cannot determine whether the audiovisual congruency effects observed here reflect a facilitatory effect of congruent sounds or an interference effect of incongruent sounds. We did have a silent condition as a baseline in Experiment 1, in which the mean detection time was comparable to an incongruent as well as a congruent condition. We speculate that prior conscious experience may have engendered a benefit of audiovisual congruency by promoting temporal integration outside of visual awareness, consistent with a causal inference account as discussed above. Future work including a baseline condition is necessary to confirm this possibility.

In sum, we provide evidence that a task-irrelevant audible sound can influence the processing of a temporally congruent visual event presented outside of visual awareness, but only when the audiovisual stimuli have been consciously experienced in advance. We also demonstrated that explicit experience of temporally incongruent audiovisual events interfered with the detection of the visual event that became congruent with the auditory event while invisible. These results suggest that previous conscious experiences of audiovisual signals may act on unconscious visual processing by modulating the extent to which they interact in the absence of visual awareness.

## CRediT authorship contribution statement

**Hyun-Woong Kim:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Minsun Park:** Writing – review & editing, Writing – original draft, Validation, Investigation. **Yune Sang Lee:** Writing – review & editing, Validation, Supervision, Resources. **Chai-Youn Kim:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgements

This work was supported by National Research Foundation of Korea grant funded by the South Korea government (NRF-2023R1A2C2007289).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.concog.2024.103709.

## References

Aller, M., Giani, A., Conrad, V., Watanabe, M., & Noppeney, U. (2015). A spatially collocated sound thrusts a flash into awareness. *Frontiers in Integrative Neuroscience, 9*. https://doi.org/10.3389/fnint.2015.00016

Alsius, A., & Munhall, K. G. (2013). Detection of audiovisual speech correspondences without visual awareness. *Psychological Science, 24*(4), 423–431. https://doi.org/10.1177/0956797612457378

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews Neuroscience, 3*(1), 13–21. https://doi.org/10.1038/nrn701

Chen, Y.-C., Huang, P.-C., Yeh, S.-L., & Spence, C. (2011). Synchronous sounds enhance visual sensitivity without reducing target uncertainty. *Seeing and Perceiving, 24*(6), 623–638. https://doi.org/10.1163/187847611X603765

Conrad, V., Bartels, A., Kleiner, M., & Noppeney, U. (2010). Audiovisual interactions in binocular rivalry. *Journal of Vision, 10*(10), 27. https://doi.org/10.1167/10.10.27

Conrad, V., Kleiner, M., Bartels, A., Hartcher O'Brien, J., Bülthoff, H. H., & Noppeney, U. (2013). Naturalistic stimulus structure determines the integration of audiovisual looming signals in binocular rivalry. *PLoS One1, 8*(8), e70710.

Cox, D., & Hong, S. W. (2015). Semantic-based crossmodal processing during visual suppression. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00722

Delong, P., Aller, M., Giani, A. S., Rohe, T., Conrad, V., Watanabe, M., & Noppeney, U. (2018). Invisible flashes alter perceived sound location. *Scientific Reports, 8*(1), 12376. https://doi.org/10.1038/s41598-018-30773-3

Faivre, N., Mudrik, L., Schwartz, N., & Koch, C. (2014). Multisensory integration in complete unawareness: Evidence from audiovisual congruency priming. *Psychological Science, 25*(11), 2006–2016. https://doi.org/10.1177/0956797614547916

Ferrari, A., & Noppeney, U. (2021). Attention controls multisensory perception via two distinct mechanisms at different levels of the cortical hierarchy. *PLOS Biology, 19*(11), e3001465.

Gau, R., & Noppeney, U. (2016). How prior expectations shape multisensory perception. *NeuroImage, 124*, 876–886. https://doi.org/10.1016/j.neuroimage.2015.09.045

Holmes, N. P., & Spence, C. (2005). Multisensory integration: Space, time and superadditivity. *Current Biology, 15*(18), R762–R764. https://doi.org/10.1016/j.cub.2005.08.058

Hong, S. W., & Shim, W. M. (2016). When audiovisual correspondence disturbs visual processing. *Experimental Brain Research, 234*(5), 1325–1332. https://doi.org/10.1007/s00221-016-4591-y

Hsiao, J.-Y., Chen, Y.-C., Spence, C., & Yeh, S.-L. (2012). Assessing the effects of audiovisual semantic congruency on the perception of a bistable figure. *Consciousness and Cognition, 21*(2), 775–787. https://doi.org/10.1016/j.concog.2012.02.001

Kayser, C., & Shams, L. (2015). Multisensory causal inference in the brain. *PLOS Biology, 13*(2), e1002075.

Kim, S., Blake, R., Lee, M., & Kim, C.-Y. (2017). Audio-visual interactions uniquely contribute to resolution of visual conflict in people possessing absolute pitch. *PLoS One1, 12*(4), e0175103.

Kohlrausch, A., Van Eijk, R., Juola, J. F., Brandt, I., & Van De Par, S. (2013). Apparent causality affects perceived simultaneity. *Attention, Perception, & Psychophysics, 75*(7), 1366–1373. https://doi.org/10.3758/s13414-013-0531-0

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One1, 2*(9), e943.

Lee, M., Blake, R., Kim, S., & Kim, C.-Y. (2015). Melodic sound enhances visual awareness of congruent musical notes, but only if you can read music. *Proceedings of the National Academy of Sciences, 112*(27), 8493–8498. https://doi.org/10.1073/pnas.1509529112

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Lewis, R., & Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *The Journal of Neuroscience, 30*(37), 12329–12339. https://doi.org/10.1523/JNEUROSCI.5745-09.2010

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic bulletin & review, 1*(4), 476–490. https://doi.org/10.3758/BF03210951

Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*(4608), 389–391. https://doi.org/10.1126/science.6867718

Moors, P., Huygelier, H., Wagemans, J., de-Wit, L., & Van Ee, R. (2015). Suppressed visual looming stimuli are not integrated with auditory looming signals: Evidence from continuous fash suppression. *I-Perception, 6*(1), 48–62. https://doi.org/10.1068/i0678

Noppeney, U. (2021). Perceptual inference, learning, and attention in a multisensory world. *Annual Review of Neuroscience, 44*(1), 449–473. https://doi.org/10.1146/annurev-neuro-100120-085519

Park, M., Blake, R., Kim, Y., & Kim, C.-Y. (2019). Congruent audio-visual stimulation during adaptation modulates the subsequently experienced visual motion aftereffect. *Scientific Reports, 9*(1), 19391. https://doi.org/10.1038/s41598-019-54894-5

Park, M., Blake, R., & Kim, C.-Y. (2024). Audiovisual interactions outside of visual awareness during motion adaptation. *Neuroscience of Consciousness, 2024*(1), 1–14. https://doi.org/10.1093/nc/niad027

Pinto, Y., Van Gaal, S., De Lange, F. P., Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision, 15*(8), 13. https://doi.org/10.1167/15.8.13

Plass, J., Guzman-Martinez, E., Ortega, L., Grabowecky, M., & Suzuki, S. (2014). Lip reading without awareness. *Psychological Science, 25*(9), 1835–1837. https://doi.org/10.1177/0956797614542132

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 385*(6614), 308.

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences, 14*(9), 425–432. https://doi.org/10.1016/j.tics.2010.07.001

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of Perceived Visual Intensity by Auditory Stimuli: A Psychophysical Analysis. *Journal of Cognitive Neuroscience, 8*(6), 497–506. https://doi.org/10.1162/jocn.1996.8.6.497

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences, 14*(9), 400–410. https://doi.org/10.1016/j.tics.2010.06.008

Tan, J.-S., & Yeh, S.-L. (2015). Audiovisual integration facilitates unconscious visual scene processing. *Journal of Experimental Psychology: Human Perception and Performance, 41*(5), 1325–1335. https://doi.org/10.1037/xhp0000074

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience, 8*(8), 1096–1101. https://doi.org/10.1038/nn1500

Van Der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance, 34*(5), 1053–1065. https://doi.org/10.1037/0096-1523.34.5.1053

Van Ee, R., Van Boxtel, J. J. A., Parker, A. L., & Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *The Journal of Neuroscience, 29*(37), 11641–11649. https://doi.org/10.1523/JNEUROSCI.0873-09.2009

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics, 72*(4), 871–884. https://doi.org/10.3758/APP.72.4.871

Vroomen, J., & Keetels, M. (2020). Perception of causality and synchrony dissociate in the audiovisual bounce-inducing effect (ABE). *Cognition, 204*, Article 104340. https://doi.org/10.1016/j.cognition.2020.104340

Yang, E., Brascamp, J., Kang, M.-S., & Blake, R. (2014). On the use of continuous flash suppression for the study of visual processing outside of awareness. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00724