Deploying attention to the target location of a pointing action modulates audiovisual processes at nontarget locations

Tristan Loria^{1,2} • Kanji Tanaka^{2,3} • Katsumi Watanabe^{1,2,4} • Luc Tremblay¹

Published online: 16 July 2020

© The Psychonomic Society, Inc. 2020

Abstract

The current study examined how the deployment of spatial attention at the onset of a pointing movement influenced audiovisual crossmodal interactions at the target of the pointing action and at nontarget locations. These interactions were quantified by measuring the susceptibility to the fission (i.e., reporting two visual flashes under one flash and two auditory beep pairings) and fusion (i.e., reporting one flash under two flashes and one beep pairing) audiovisual illusions. At movement onset, unimodal, or auditory and visual bimodal stimuli were either presented at the target of the pointing action or in an adjacent, nontarget location. In Experiment 1, perceptual accuracy within the unimodal and bimodal conditions was lower in the nontarget relative to the target condition. The fission illusion was uninfluenced by target condition. However, the fusion illusion was more likely to be reported at the target relative to the nontarget location. In Experiment 2, the stimuli from Experiment 1 were further presented at a location near where the eyes were fixated (i.e., congruent condition), where the hand was aiming (i.e., target), or in a location where neither the eyes were fixated nor the hand was aiming. The results yielded the greatest susceptibility to the fusion illusion when the visual location and movement end points were congruent relative to when either movement or fixation was incongruent. Although attention may facilitate the processing of unisensory and multisensory cues in general, attention might have the strongest influence on the audiovisual integration mechanisms that underlie the sound-induced fusion illusion.

Keywords Vision · Audition · Attention · Multisensory · Pointing

Multisensory perception occurs via the combination and integration of signals from the various sensory channels (i.e., vision and audition in the context of the present study). Crossmodal interactions, in which information presented in one sensory modality influences perception in an alternate modality, is one mechanism underlying multisensory perception (e.g., Spence & Driver, 1997). Once considered neurophysiological in nature, recent behavioral experiments have examined how multisensory perception may be influenced by cognitive or perceptual processes like attention (i.e., for a

review, see van Atteveldt, Murray, Thut, & Schroeder, 2014). Although our understanding of attention is admittedly limited (see Hommel et al., 2019), attention has often been characterized as a “spotlight” in which information falling within the “spotlight” receives a processing advantage relative to information falling outside (e.g., Treue, 2001). Because attention is a limited capacity resource (i.e., Lavie, 1995; Treisman & Gelade, 1980), attentional deployment must be flexible (e.g., Kahneman, 1973) to process salient information that is relevant to an individual’s goals (i.e., Broadbent, 1958; Cherry & Taylor, 1954; Folk, Remington, & Johnston, 1992). Much like multisensory integration, attention is a mechanism that facilitates perception of the environment (e.g., Posner, Snyder, & Davidson, 1980). Interestingly, studies examining how attention interacts with multisensory perception have reached varied conclusions (i.e., for a review, see Macaluso et al., 2016).

Directing attention to multisensory cues has typically been used to examine the relation between attention and multisensory integration. For example, Busse, Roberts, Crist, Weissman, and Woldorff (2005) employed a task in which participants were instructed to attend or ignore a visual stimulus while brain activity was recorded using functional magnetic resonance imaging (fMRI). When a task-irrelevant

✉ Luc Tremblay
luc.tremblay@utoronto.ca

¹ Centre for Motor Control, Faculty of Kinesiology and Physical Education, University of Toronto, Toronto, Canada

² Faculty of Science and Engineering, Waseda University, Tokyo, Japan

³ Faculty of Arts and Science, Kyushu University, Fukuoka, Japan

⁴ Art & Design, University of New South Wales, Sydney, Australia

auditory tone was presented, a response enhancement localized to the auditory cortex was observed only when the task-relevant visual stimulus was attended, suggesting that attention within one modality can enhance processing in an alternate unattended modality. Additional work (e.g., Donohue, Green, & Woldorff, 2015) has also shown that attending to the spatial or temporal features of a multisensory event can either broaden or narrow the temporal window of integration (i.e., a mechanism that is considered fundamental to multisensory integration; see Colonius & Diederich, 2004). These facilitation-like effects can be interpreted as evidence that multisensory perception is potentially enhanced by directing attention towards multisensory stimuli in a top-down fashion. However, alternate work has shown that multisensory perception is not facilitated or inhibited when attention is constrained within a dual-task paradigm, or directed away from the spatial location where a multisensory cue was presented (i.e., Bertelson, Vroomen, de Gelder & Driver, 2000; Helbig & Ernst, 2008; Vroomen, Bertelson, & de Gelder, 2001). These additional studies can suggest that multisensory perception is carried out in a bottom-up fashion. One potential way to disentangle the conditions in which attention is most likely to influence multisensory integration may be to employ additional behavioral paradigms that typically facilitate sensory processing at attended relative to unattended locations.

Attention is often deployed via the eyes to the location where an upcoming goal-directed action will be executed (i.e., oculomotor attention). In a study conducted by Deubel and Schneider (1996), participants were cued to plan a saccade to a location of a masked target. Sixty milliseconds following the go signal, the target (i.e., as revealed by partially removing the mask stimuli) could be presented at the same or at a different location as the saccade target, while the remaining masks were removed displaying distractor stimuli. Participants provided a button-press response to indicate the position where the target had been presented. When the target had been presented at the location of the intended saccade, response accuracy was nearly perfect. However, when the target was presented outside of the saccade location, performance was not statistically different from chance. Deubel and Schneider (1996) reasoned that because the target was presented prior to the saccade, the improvement in target detection was driven by the deployment of oculomotor attention to the location of the upcoming saccade which facilitated information processing at the to-be-foveated location. A similar facilitation has been reported using an upper-limb pointing paradigm.

Information presented in proximity to the hand's path is often attended more so than locations further away from the hand. In their seminal study, Tipper, Lortie, and Baylis, (1992) had participants perform pointing movements to target locations arranged in a three-by-three matrix. Participants performed movements to targets (i.e., indicated by a red light)

while ignoring distractors (i.e., indicated by a yellow light). In Experiment 1, movements originated at the bottom of the target matrix (i.e., close to the participant's body). When participants reached to targets in the middle row, distractors within the first and the middle row caused the greatest amount of interference on total response time. In Experiment 2, movements originated from the top of the aiming console (i.e., away from the participant's body). With this configuration, distractors presented in the middle and top row (i.e., the row closest to the start position) caused the greatest amount of interference. Tipper et al. (1992) surmised that because most interactions with the environment occur via the motor system, attention must have a robust action-based component. As a result, manipulating attention from a goal-directed action perspective (i.e., termed here as action-based attention) may provide a novel lens to examine the hypotheses of whether attention influences multisensory perception.

Audiovisual sound-induced flash illusions have previously been used to quantify multisensory perception and crossmodal interactions, but only applied to a limited extent in studies involving reaching (cf. Tremblay & Nguyen, 2010). The fission illusion occurs when a single flash presented alongside two or more beeps yields the perception of two or more flashes (e.g., Shams, Kamitani, & Shimojo, 2000). The fusion illusion occurs when two flashes paired with a single beep yield the perception of a single flash (e.g., Andersen, Tiipana, & Sams, 2004; Shams & Kim, 2010). Interestingly, the fission illusion occurs more frequently in the periphery than at the fovea (i.e., 5-degree eccentricity; see Chen, Maurer, Lewis, Spence, & Shore, 2017), whereas fusion is less or unaffected by visual eccentricity (i.e., Tremblay et al., 2007). One could surmise that the effect of eccentricity could be related to the distribution of rods and cones on the retina, thus influencing the likelihood of perceiving one versus two flashes. As originally described by Osterberg (1935), central vision is predominately cone-based, and thus is dominated by parvocellular processes, whereas peripheral vision is dominated by rods (i.e., magnocellular processes). Given this difference in basic visual processing pathways, it can be hypothesized that alterations in the susceptibility to the illusions in central versus peripheral vision may reflect a response rather than perceptual bias.¹ Although the influence of basic visual processes on the audiovisual illusions is beyond the scope of the present study (i.e.,

¹ Our group has further investigated the relationship between eccentricity, attention, and multisensory perception. The studies in question involved nearly identical procedures as those reported in the present study. Briefly, we have found that presenting the fusion illusion further into the periphery (i.e., 30 degrees) yields comparable susceptibility with when the illusion was presented at the target of the movement. In addition, the modulation of the fusion illusion at central versus peripheral vision occurred only when a goal-directed action was executed. These results could suggest a perceptual, rather than response bias as the most likely interpretation of altered illusion susceptibility in relation to attention (i.e., see published abstracts: Loria, Hajj, Tanaka, Watanabe, & Tremblay, 2018, 2019).

but see General Discussion), the previously reported dissociation, with regard to the spatial location in which the illusions were presented, may at least suggest that the fission and fusion illusions occur because of distinct underlying crossmodal processes, rather than a common crossmodal process (i.e., Chen et al., 2017; Mishra, Martinez, & Hillyard, 2010). Such a hypothesis was investigated in the current study using goal-directed reaching.

Employing a goal-directed reaching task can provide an ecologically valid method to examine how attentional deployment via the eyes and hands reciprocally influence multisensory perception. When reading a newspaper, for example, oculomotor attention is dedicated to the words on the page. While attention is engaged with the newspaper, an individual may also reach to the remembered position of a cup of coffee without disengaging the eyes from the page. Given the success of such behavior in daily life, attention can effectively be divided across oculomotor and action-based reaching tasks. Considering this example, the current study examined whether deploying attention at the onset of a pointing action (i.e., Experiment 1), in relation to eye position and target location (i.e., Experiment 2), would facilitate or inhibit the individual pathways underlying both sound-induced flash illusions.

Participants executed pointing movements to a central square that was flanked on either side by additional squares. At movement onset, visual flashes (F) were presented within one of the squares, and auditory beeps (B) were presented via speakers. The visual flashes were predominately presented within the central square (i.e., which was always the target of the pointing action), but were also presented within the nontarget squares (i.e., the nontarget flanking squares). Upon movement completion, participants judged how many flashes had been presented. Given the previous reports of facilitation of information processing at the terminal position of an upcoming saccade (i.e., Deubel & Schneider, 1996), as well as the deployment of attention during a pointing movement (i.e., Tipper et al., 1992), it was hypothesized that attention would be deployed to the terminal position of the pointing movement. Directing attention to the target was expected

to facilitate unimodal (i.e., 1F0B; 2F0B) and multisensory bimodal (i.e., 1F1B) judgements relative to the nontarget locations. Such a finding would provide additional evidence that attention influences multisensory integration in a top-down fashion.

Directing attention during the reaching movement was also expected to influence the sound-induced flash illusions. One could predict that, if the illusions share a common mechanism, the susceptibility to both the illusions would be increased at the target relative to the nontarget locations (i.e., as indicated by lower accuracy scores). However, the illusions may be the product of unique audiovisual mechanisms (e.g., Chen et al., 2017; see also Mishra, Martinez, & Hillyard, 2010; Mishra, Martinez, Sejnowski, & Hillyard, 2007). As a result, it may be additionally hypothesized that the influence of attention could be discrepant between the two illusions. Such a finding would extend the current understanding of the conditions under which the illusions are most likely to occur to encompass dynamic behavior (i.e., goal-directed action) that has yet to be thoroughly examined.

Experiment 1

Methods

Participants

The criteria used for attaining an adequate sample size was derived from previous work on the fission illusion (e.g., Bolognini, Rossetti, Casati, Mancini, & Vallar, 2011; Shams et al., 2000). Thirteen individuals from the student population at Waseda University completed the experiment. Of these 13 individuals, one participant was excluded for having 40% of trials with movement end points along the medial-lateral axis greater than two standard deviations from the group mean (see Table 1). In addition, another participant was excluded for having an average movement end point distribution along the sagittal axis greater than two standard deviations from the group mean

Table 1 The perceived number of flashes are reported in relation to the number of flashes and beeps physically presented per trial. In addition, results for the kinematic variables are shown. The associated units of measurement for the kinematic variables are reported for each individual variable

Target condition	Number of flashes, number of beeps					Kinematic variables			
	1, 0	1, 1	1, 2	2, 0	2, 1	Movement time (s)	Decision time (s)	Medial-lateral error (mm)	Sagittal error (mm)
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
Target	1.18 (.09)	1.16 (.13)	2.06 (.29)	2.22 (.19)	1.35 (.32)	1.01 (.48)	.50 (.42)	3.42 (12.59)	11.05 (17.49)
Nontarget	1.20 (.42)	1.48 (.38)	2.18 (.21)	2.52 (.23)	2.19 (.44)	1.02 (.38)	.44 (.06)	3.92 (8.87)	11.85 (11.78)

(see Table 1). As such, data from 11 participants (four males) with a mean age of 21.3 years were included in formal data analysis. All participants self-reported to be right-hand dominant, with normal or corrected-to-normal vision. The experiment was approved by the Institutional Review Board at Waseda University. Participants provided informed written consent prior to completing the experiment and were compensated ¥1,000.

Apparatus

Participants sat in front of a horizontally oriented touch-screen monitor (24-inch SHARP, LL-S242A-W, Sharp Corporation, Osaka, Japan) with a resolution of $1,920 \times 1,080$ pixels and a refresh rate of 60 Hz. Testing took place within a sound-attenuated chamber with a background sound level of 32 dB. Movements of the head were restricted via a chin rest, ensuring that the visual distance of 35 centimeters (cm) between the eyes and the computer monitor was held constant for all experimental trials. Such a distance from the monitor was required to ensure participants could properly reach the targets on the screen. Auditory stimuli were presented via two speakers (BOSE, Companion® 20 multimedia speaker system, Bose Corporation, Framingham, MA, USA) positioned on either side of the monitor. The experiment was controlled using a customized MATLAB script (The MathWorks, Natick, MA, USA) using the Psychtoolbox extension (Brainard, 1997).

Stimuli

Three squares, each $1 \text{ cm} \times 1 \text{ cm}$, were displayed on the computer monitor. The home position was a 1 cm^2 white circle positioned 1 cm above the bottom edge of the monitor.

On every trial, participants made judgments regarding the number of times a white disk was presented. The white disk was 1 cm^2 (.06 lux: REED LX-102 Light Meter, REED Instruments, Wilmington, NC, USA) and was presented on each trial. The disk flashed once or twice and could be accompanied by auditory beeps. Varying the number of flashes and beeps yielded the following flash (F) and beep (B) combinations: 1F0B, 2F0B, and 1F1B, 1F2B (i.e., fission illusion), and 2F1B (i.e., fusion illusion). In an effort to reduce any fatigue associated with repeatedly reaching vertically against the forces of gravity, and to ensure that the experiment was completed in 60 minutes, participants completed 280 trials in total. Of the total trials performed, 40 of each sensory condition were presented at the target of the pointing action, with an additional 16 trials of each condition presented at nontarget locations. The center-to-center distance between the target and nontarget locations was 12.5 cm. The auditory beeps had a 5-ms linear ramps at onset and offset and were presented at maximum intensity (i.e., 72 dB) for 7 ms at frequency of 3.5 kHz (see Fig. 1).

Procedure

Participants were instructed to place their finger on the home position to initiate every trial. When the finger was placed on the home position, a fixation cross would appear above the central rectangle. Participants were instructed to fixate on the fixation cross above the central square for the remainder of the trial. The central rectangle ($1 \text{ cm} \times 1 \text{ cm}$) would then flash (i.e., 50 ms offset), indicating the target of the pointing action. Such a short target presentation duration was employed to ensure participants were fixating on the fixation cross. After the central rectangle flash, participants began their movement vertically towards the central target with the goal of

Fig. 1 Depiction of the experimental task and target conditions. Participants initiated each trial by placing their finger on the home position. Following the target cue, participants moved their finger vertically towards the central target square. When the finger lost contact

with the home position, the experimental stimuli were presented at target or nontarget locations. The timing of the flashes and beeps for the fission and fusion illusions are shown

terminating their movement at the lower border of the rectangle (i.e., to avoid the potential for the fingertip to occlude the presentation of the disk). Participants always performed movements towards the central target.

The stimuli (i.e., unimodal, bimodal congruent, or bimodal incongruent) were triggered at movement onset (i.e., when the finger released contact with the home position). Target trials occurred when the stimuli were presented within the central target. Nontarget trials occurred when the stimuli were presented in one of the two flanking targets (see Fig. 1). Following movement termination, a message displaying “How many visual flashes did you see?” was presented, and participants pressed the corresponding box on screen (i.e., between one, two, and three flashes). Trials were self-paced, and participants placed their finger on the home position to initiate the subsequent trial.

Prior to completing the experimental trials described above, a familiarization block was completed. The familiarization block was divided into two phases. Both phases consisted of 10 trials in which the specific flash/beep stimuli were presented randomly. In the initial block, participants were instructed to focus on performing the reaching movement while keeping their eyes on the fixation cross. The second familiarization phase focused on judgment accuracy. Participants were instructed to withhold from entering their response manually and instead report it verbally to the experimenter. The experimenter, who was trained at discerning the flashes presented, provided feedback if the participant’s response was incorrect. If participants were less than 50% accurate (i.e., as noted by the experimenter), this second familiarization phase was completed again. Two participants required additional accuracy-based familiarization trials, with no participants requiring an additional set of reaching familiarization trials.

Data analysis

Movement time was defined as the time, in seconds, from the release of the home position to when the finger made contact again with the touch screen (i.e., at the target location). Decision time was the time in seconds for participants to provide a judgement on the number of flashes. Medial-lateral error was defined as the movement end point along the medial-lateral axis (i.e., left and right in the pointing environment), subtracted from the bottom edge of the central target position. Sagittal error was defined as the movement end point along the vertical axis, subtracted from the bottom edge of the central target location. All kinematic related variables discussed above were submitted to paired-samples *t* tests, with target condition (i.e., target or nontarget) as the independent variable.

Flash accuracy was the main dependent variable and was calculated by comparing the reported number of flashes to the

number of physically presented flashes per trial (see also Chen et al., 2017). A correct response occurred when the reported number of flashes matched the physical number of flashes on that trial. For flash accuracy, average responses were collapsed across nontarget conditions (i.e., left-of-center and right-of-center squares). The unimodal control conditions (i.e., 1F0B; 2F0B) were further collapsed into a single unimodal sensory cue condition. Accuracy scores were submitted to a 4 (sensory condition: unimodal vs. bimodal vs. fission vs. fusion) $\times$ 2 (target condition: target vs. nontarget) repeated-measures analysis of variance (ANOVA). Tukey’s honestly significant difference (HSD) tests were conducted for the significant effects of interest. Lastly, in cases of significant violations of sphericity (i.e., as measured by Mauchly’s test of sphericity), Greenhouse–Geisser-corrected degrees of freedom were reported to the nearest decimal.

Results

A detailed summary of all dependent variables can be found in Table 1. Movement time did not significantly differ between target and nontarget conditions, $t(10) = 1.25$, $p = .242$, 95% CI $[-.056, .016]$. Decision time also did not differ between target and nontarget conditions, $t(10) = 1.08$, $p = .304$, 95% CI $[-.638, .022]$. Medial-lateral error did not differ between the target and nontarget conditions, $t(10) = .446$, $p = .67$, 95% CI $[-2.82, 1.88]$, nor did sagittal error, $t(10) = .442$, $p = .67$, 95% CI $[-4.85, 3.25]$. Overall, movements and the timing of the perceptual judgments regarding the number of flashes did not significantly differ between the target and nontarget conditions.

The results for flash perception accuracy revealed a main effect of sensory cue condition, $F(1.9, 19.2) = 85.63$, $p < .001$, $\eta_p^2 = .90$, and a main effect of target condition, $F(1, 10) = 8.84$, $p = .014$, $\eta_p^2 = .47$. These effects were superseded by a significant Sensory Cue Condition $\times$ Target Condition interaction, $F(3, 30) = 15.99$, $p < .001$, $\eta_p^2 = .62$.² Post hoc tests (HSD = 15%) revealed that accuracy was significantly greater when the secondary flash/beep cues were presented at the target location ($M = 81\%$, $SD = 6$) relative to the nontarget locations ($M = 65\%$, $SD = 15$) in the unimodal condition. Within the bimodal condition, accuracy was greater at the target ($M = 82\%$, $SD = 8$) compared with the nontarget locations ($M = 60\%$, $SD = 26$). For the results describe below regarding the sound-induced flash illusions, it is important to

² As an exploratory analysis, the perceptual accuracy analysis was carried out including the two participants who had been excluded from formal data analysis. Including these participants revealed the same main effect of sensory condition and Sensory Condition $\times$ Target Condition interaction. No significant effect of target condition was revealed. However, note that these participants were excluded based on movement end points along the medial-lateral and sagittal axis, thus suggesting that these two participants were not aiming accurately at the center target.

note that lower accuracy indicates greater illusion susceptibility. Accuracy for the fission illusion did not significantly differ between target ($M = 6\%$, $SD = 8$) and nontarget ($M = 2\%$, $SD = 3$) locations. However, when presented with the flash/beep cue combinations associated with the fusion illusion, participants were more accurate at the nontarget ($M = 49\%$, $SD = 19$) than at the target location ($M = 27\%$, $SD = 22$; see Fig. 2).

Discussion

Participants completed pointing movements to a central target. Unimodal, bimodal, fission, and fusion flash/beep stimuli were presented at target and nontarget locations. Within the unimodal and bimodal conditions, judgments regarding the number of flashes presented were most accurate at the target relative to the nontarget locations. However, the target condition yielded no significant influence on judgment accuracy of the fission illusion, while the fusion illusion was more likely to be perceived at the target location relative to the nontarget locations. Below, these findings were considered in terms of the influence of spatial attention on unimodal, bimodal, and sound-induced flash illusion perception.

In the absence of audiovisual conflicts, participants were more accurate in the target relative to the nontarget condition. The effect of target condition supplements previous reports in which luminance changes or target presentation within an expected, or to-be-foveated location is most efficiently and accurately processed (i.e., Deubel & Schneider, 1996; Downing, 1988; Eriksen & Yeh, 1985; Hoffman & Nelson, 1981; Müller & Findlay, 1987; Posner, Snyder, & Davidson, 1980). The effects observed for the unimodal, bimodal, and

sound-induced flash illusion conditions suggest that spatial attention was effectively deployed to the target of the pointing movement. As a result, the findings of Experiment 1 can support the hypothesis that attention influences multisensory integration, and specifically that attention increased the susceptibility to the fusion illusion relative to when the illusion was unattended.

Although the susceptibility to the fission illusion was comparable across target conditions, the fusion illusion was more likely to be perceived when presented at the target of the pointing action relative to the nontarget locations. As such, the implications of the current findings for top-down and bottom-up modulation of audiovisual crossmodal interactions was mixed. Regarding fission, the lack of influence of target condition may suggest that the resulting percept of two flashes does not occur following the influence of attention, meaning fission may be predominantly bottom up and not influenced by attention (i.e., Bertelson et al., 2000; Helbig & Ernst, 2008; Vroomen et al., 2001). However, an alternate explanation could be that the fission results reported here reflect a ceiling effect due to the considerable susceptibility to the illusion. An attempt at disentangling this alternate explanation was made in Experiment 2. The results for fusion indicate that the propensity to see the illusion potentially relies on attending to the location where the stimuli were presented (i.e., aiming at the target), supporting a top-down contribution to audiovisual multisensory perception (i.e., Busse et al., 2005; Donohue et al., 2015; Talsma, 2015; Talsma, Doty, & Woldorff, 2007). The theoretical implications of the dissociation reported here are described below in the General Discussion. Prior to that discussion, however, a limitation of Experiment 1 was examined.

One could argue that a limitation of Experiment 1 was that the spatial position where the illusion stimuli were presented was not equivalent in terms of visual fixation in the target versus nontarget condition. For example, when presented at the target of the pointing action, the stimuli were below the fixation point. In contrast, when presented in the nontarget locations, the stimuli were in the visual periphery (i.e., 20-degree eccentricity). As mentioned in the Introduction, presenting the illusory stimuli in the periphery has been previously shown to increase susceptibility to the fission illusion, and decrease susceptibility to the fusion illusion (i.e., Chen et al., 2017; Tremblay et al., 2007). As such, the current findings cannot be solely attributed to the influence of top-down and bottom-up crossmodal processes, but could relate to differences of visual information presented on portions of the retina. To address this issue, Experiment 2 was conducted.

In Experiment 2, the spatial position of fixation was manipulated relative to the presentation of the flash/beep stimuli and the pointing action. Participants completed aiming movements to the central target, while fixation was directed towards

Fig. 2 Response accuracy data are shown. Error bars denote between subject standard error of the mean. The asterisks indicate significant differences as revealed by the post hoc analysis of the Sensory Cue Condition $\times$ Target Condition interaction. Participants overall were highly susceptible to the fission illusion at both target and nontarget locations. However, susceptibility to the fusion illusion was significantly greater at the target than the nontarget location

either the left square, the center square (i.e., the target of the pointing action), or the right square. Manipulating visual fixation in this way yielded conditions in which the participants were (1) aiming to and fixating the same location, (2) aiming to but not fixating the same location, (3) fixating at one location but aiming to an alternate location, or (4) neither fixating nor aiming to the same spatial location. It was hypothesized that aligning visual and spatial attention (i.e., aiming and fixating at the same location) would yield the greatest advantage for perceptual accuracy in the unimodal and bimodal conditions, while yielding the lowest accuracy (i.e., indicating illusion susceptibility) for the fusion illusion specifically. If confirmed, such effects could provide further evidence for the influence of attention on multisensory perception.

Experiment 2

Methods

All experimental details were identical to Experiment 1, except for the following details.

Participants

Twelve participants (four male) from the University of Toronto student population completed Experiment 2. The average age was 23.2 years, with a standard deviation of 6.1 years. Participant's handedness was confirmed using a modified version of a hand dominance test (i.e., Oldfield, 1971). The Research Ethics Board at the University of Toronto approved the experimental protocol. Participants were compensated \$10 CAD.

Apparatus

Participants completed the experiment on a 3M Touch Systems Inc. touch-screen monitor (model number: 98-0003-3598-8, Maplewood, MN, USA). The resolution of the screen was $1,680 \times 1,050$, and the refresh rate was 60 Hz. Auditory stimuli were presented via Altec Lansing Series 100 speakers (Altec Lansing, New York City, NY, USA) positioned on either side of the monitor.

Stimuli

Participants completed three blocks of 140 trials (i.e., 420 in total). Blocking the trials was deemed critical in the absence of eye position monitoring equipment. Within each block, participants were directed via onscreen instructions to fixate on the square indicated for that block of trials. Participants completed 140 trials each while fixating on the left, center, and right squares.

The order of fixation was counterbalanced across participants. Of the 140 trials performed in each block, 10 of each sensory condition were presented at the target of the pointing action plus an additional 10 catch 0F0B trials. The catch trials were included due to the additional fixation conditions employed in Experiment 2, which lengthened the testing time. To ensure participants were still attentive, the addition of catch trials to Experiment 2 was deemed critical. In addition, 10 trials of each sensory condition were presented in the nontarget locations in addition to the ten catch 0F0B trials. As such, it was equally probable that the stimuli occurred at the target of the pointing action or within a nontarget location.

Procedure

Participants still performed aiming movements at the central target square only. Following movement termination, when the “How many visual flashes did you see?” message was presented, participants had the opportunity to respond between zero, one, and two flashes. The familiarization blocks were completed prior to each experimental block (i.e., three times in total) to adapt participants to the novel eye positions employed in each block. Three participants required additional accuracy-based familiarization trials across all experimental blocks.

Data analysis

Responses were grouped according to the target, with congruency referring to the location of the sensory cues. The movement congruent, fixation congruent (MCFC) condition was the mean of all trials in which the sensory cues were presented at the target (i.e., center) location. The movement congruent, fixation incongruent (MCFI) was the mean of all trials in which the sensory cues occurred at the central target position and the eyes were directed to the left or right target. The movement incongruent, fixation congruent (MIFC) condition was the mean of all trials where the sensory cues were presented in the left or right target locations where the eyes were fixated. Lastly, the movement incongruent, fixation incongruent (MIFI) condition comprised all trials where the stimuli were presented in neither of the target or fixation locations (see Fig. 3). For the analyses, 1F0B and 2F0B trials were collapsed into the unimodal condition. In addition, 1F1B and 2F2B trials were collapsed into the bimodal condition. The dependent variables were: movement time, decision time, medial-lateral error, sagittal error, and perceptual accuracy scores, which were submitted to separate 4 (sensory cue condition: unimodal vs. bimodal vs. fission vs. fusion) $\times$ 4 (congruency: MCFC vs.

Fig. 3 Depiction of the levels of target shown in relation to fixation locations. Congruency relates to where the stimuli were presented. Participants always aimed to the center square. In the text, the left square refers to the left-of-center square, the center square refers to the

central square, and the right square refers to the right-of-center square. Note that the fixation cross is only present in the figure to represent where the eyes were positioned and was not physically present during the experiment

MCFI vs. MIFC vs. MIFI) repeated-measures ANOVAs. All post hoc tests were carried out using Tukey's HSD test.

Results

Kinematic data can be found in Table 2. Analysis of movement times yielded a main effect of sensory cue condition, $F(1.4, 14.9) = 6.65$, $p = .015$, $\eta_p^2 = .38$, but no effect of congruency or Sensory Cue Condition $\times$ Congruency interaction (i.e., all p values $> .13$). Post hoc tests (HSD = 8 ms) for the main effect of sensory cue condition revealed that the unimodal ($M = 464$ ms, $SD = 17$) and bimodal ($M = 469$ ms, $SD = 16$) conditions did not significantly differ from each other, but yielded significantly longer movement times than the fission condition ($M = 450$ ms, $SD = 15$). In contrast, the unimodal, bimodal, and fission condition movement times were all significantly shorter than the fusion illusion movement times ($M = 509$ ms, $SD = 9$). Decision time yielded a main effect of sensory cue condition, $F(3, 33) = 3.30$, $p = .033$, $\eta_p^2 = .52$, but no effect of congruency or Sensory Cue Condition $\times$ Congruency interaction (i.e., all p values $> .10$).

Post hoc tests (HSD = 18 ms) for the effect of sensory cue condition revealed that decision times were longer in the fission ($M = 524$ ms, $SD = 20$) and fusion ($M = 531$ ms, $SD = 18$) relative to the bimodal ($M = 500$ ms, $SD = 17$) condition. Medial-lateral error yielded a main effect of sensory cue condition, $F(3, 33) = 7.51$, $p = .001$, $\eta_p^2 = .41$, no main effect of congruency, $F(1.3, 13.4) = .68$, $p = .50$, $\eta_p^2 = .06$, and a significant Sensory Cue Condition $\times$ Congruency interaction, $F(3.1, 33.7) = 3.4$, $p = .30$, $\eta_p^2 = .24$. Note that larger error values indicate a greater bias towards the right side of the reaching environment. However, post hoc tests performed for the sensory cue condition and congruency interaction (HSD = 4.12 mm) failed to yield differences. Analysis of sagittal error revealed a main effect of sensory cue condition, $F(3, 33) = 3.82$, $p = .02$, $\eta_p^2 = .26$, and a main effect of congruency, $F(1.3, 13.8) = 7.52$, $p = .01$, $\eta_p^2 = .40$, but no Sensory Cue Condition $\times$ Congruency interaction, $F(2.9, 32.9) = 1.3$, $p = .31$, $\eta_p^2 = .10$. Note that for sagittal error, values closer to zero reflect movements that terminated closer to the target's location. Post hoc tests performed for the effect of sensory cue condition (HSD = 3.22 mm) failed to reveal

Table 2 Movement related (i.e., kinematic) variables are shown. Fixation location refers to the square that participants were fixating on that block of trials. The square in which the cues were presented refers to

the square in which the experimental stimuli were presented, irrespective of the target of the pointing action. Note that ML error refers to medial-lateral error

Kinematic variable												
Movement time (s)				Decision time (s)			ML error (mm)			Sagittal error (mm)		
<i>M (SD)</i>				<i>M (SD)</i>			<i>M (SD)</i>			<i>M (SD)</i>		
Square in which the flash/beep cue was presented												
Fixation location	Left	Center	Right	Left	Center	Right	Left	Center	Right	Left	Center	Right
Left	.54 (.40)	.54 (.41)	.53 (.36)	.52 (.20)	.54 (.23)	.57 (.22)	1.1 (6)	1.4 (5.3)	.27 (6.3)	13.3 (9.8)	10.6 (9.9)	11.9 (9.7)
Center	.51 (.30)	.52 (.32)	.50 (.29)	.50 (.19)	.47 (.16)	.49 (.18)	3.1 (4.2)	3.4 (3.7)	3.7 (4.6)	10.5 (5.1)	8.6 (5.4)	10.6 (6.9)
Right	.52 (.30)	.52 (.29)	.52 (.31)	.51 (.19)	.52 (.19)	.57 (.22)	7.6 (6.5)	5.5 (6.8)	5.9 (6.5)	11.4 (10.2)	8.1 (8.5)	11.6 (10.1)

significant differences between the conditions. Lastly, post hoc tests (HSD = 7.3 mm) performed for the main effect of congruency also failed to reveal significant differences between conditions.

The reported number of flashes per target condition can be found in Table 3. The analysis of flash accuracy yielded a main effect of sensory cue condition, $F(1.8, 20.3) = 28.40$, $p < .001$, $\eta_p^2 = .71$, a main effect of congruency, $F(3, 33) = 6.11$, $p = .002$, $\eta_p^2 = 0.36$, and a significant Sensory Cue Condition $\times$ Congruency interaction, $F(9, 99) = 14.80$, $p < .001$, $\eta_p^2 = .57$. Post hoc tests performed for the interaction (HSD = 15%) revealed that within the unimodal condition, accuracy was greater in the MCFC ($M = 68\%$, $SD = 17$), MCFI ($M = 60\%$, $SD = 7$), and MIFC ($M = 62\%$, $SD = 14$) conditions, compared with the MIFI ($M = 41\%$, $SD = 6$) condition. In the bimodal condition, accuracy was significantly greater in the MCFC ($M = 66\%$, $SD = 16$) and MCFI ($M = 67\%$, $SD = 8$), and MIFC ($M = 62\%$, $SD = 16$) conditions relative to the MIFI ($M = 41\%$, $SD = 6$). Note that for the sound-induced illusions, lower accuracy reflects greater illusion susceptibility. Analysis of fission accuracy failed to yield any significant differences between sensory and congruency

conditions (i.e., largest mean difference = 9%). Lastly, accuracy in the fusion condition revealed the greatest level of accuracy was achieved in the MIFI ($M = 46\%$, $SD = 18$) relative to the MCFC ($M = 11\%$, $SD = 16$), MCFI ($M = 29\%$, $SD = 14$), and MIFC ($M = 29\%$, $SD = 16$) conditions. The lowest accuracy observed was found for the MCFC condition relative to the MCFI, and MIFC conditions. Accuracy findings are shown in Fig. 4.

Discussion

Experiment 2 was conducted to examine whether visual attention arising from visual fixation and visual fixation combined with goal-directed pointing influenced the processing of unimodal, audiovisual bimodal, and audiovisual illusion stimuli. Experiment 2 revealed that the accuracy of unimodal and bimodal judgments was influenced by visual fixation, while engaging in a goal-directed pointing movement significantly influenced susceptibility to the fusion illusion. Indeed, the fusion illusion was more likely to be perceived when presented where the hand and eyes were aligned (i.e., MCFC condition) relative to the MCFI condition and MIFC condition. As a result, the current findings regarding the susceptibility to the illusions may suggest that deploying attention to the target of a

Table 3 Mean number of reported flashes are presented in relation to target condition and number of flashes and beeps physically presented

	Number of flashes, number of beeps					
	Unimodal 1, 0	Bimodal 1, 1	Fission 1, 2	Bimodal 2, 0	Fusion 2, 1	Catch 0, 0
Target condition	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
MCFC	1.37 (.28)	1.31 (.18)	1.69 (.10)	1.69 (.10)	1.10 (.15)	.03 (.08)
MCFI	1.40 (.20)	1.33 (.19)	1.65 (.17)	1.60 (.13)	1.31 (.15)	.06 (.10)
MIFC	1.46 (.30)	1.40 (.24)	1.69 (.11)	1.63 (.23)	1.28 (.17)	.02 (.07)
MIFI	1.58 (.29)	1.54 (.24)	1.65 (.21)	1.61 (.24)	1.57 (.21)	.03 (.07)

Fig. 4 Perceptual accuracy findings are shown as a function of sensory cue and target location condition. Error bars denote standard error of the mean. The asterisks indicate significant differences between the

conditions as revealed by the post hoc analysis conducted on the Sensory Cue Condition $\times$ Congruency interaction

pointing action further modulated the illusory percepts in addition to the visual eccentricity effects previously reported (e.g., Chen et al., 2017). Below, the results for the kinematic analyses and perceptual accuracy findings were discussed.

The movement time and decision-time findings require further consideration. Interestingly, the fusion illusion revealed longer movement times relative to the unimodal and bimodal conditions, which were in turn longer than those in the fission conditions. The longer movement times for fusion relative to the unimodal and bimodal conditions may suggest that additional processing resources (i.e., perhaps those involved in attending to sensory cues) were engaged over and above the processing resources dedicated to controlling the action. However, such a hypothesis is indeed speculative and would require further evidence. Regarding decision time, it would appear that the illusion conditions (i.e., incongruent pairing of auditory and visual cues) increased decision times regarding the number of flashes presented relative to the bimodal condition (i.e., congruent pairing of auditory and visual cues). Such an increase in time spent preparing a response could be indicative of a more challenging perceptual process (e.g., Huang & Pashler, 2005; Palmer, Nasman, & Wilson, 1994), which could have affected the ongoing motor response. When examining the effects reported for movement time and decision time, the data would certainly support an attention-based interpretation of the data relative to a fixation hypothesis. The results for perceptual accuracy also fit well with an attention-based perspective.

Reports of fusion were enhanced and decreased respectively as a function of congruency condition. Specifically, the fusion illusion was more likely to be perceived in the MCFC

relative to the MCFI and MIFC conditions. That is, when fixation and the movement's end point were aligned, participants were more likely to perceive the fusion illusion relative to when the illusion was presented at the movement's end point, but fixation was directed elsewhere, or fixation was located where the illusion was presented, but movements were executed elsewhere. As such, the results of Experiment 2 may be taken as preliminary evidence that susceptibility to the fusion illusion is not strictly influenced by visual fixation alone (i.e., Chen et al., 2017; Tremblay et al., 2007), but also by deploying attention at the onset of a pointing action, and further by the alignment between action-based and oculomotor attention. Such an interpretation requires being considered along with the results of Experiment 1.

General discussion

Experiment 1 employed a goal-directed reaching paradigm to determine the influence of attention on multisensory integration and crossmodal interactions. Participants performed a reaching action to a central target where unimodal and bimodal stimuli were mostly presented. On some trials, these stimuli were presented at a location outside of the reaching target (i.e., nontarget location). When presented at the nontarget location, the accuracy in judging the number of flashes presented was reduced relative to the target location for unimodal and bimodal trials. The susceptibility to the fission illusion did not differ between target conditions, whereas the fusion illusion was more likely to be reported at the nontarget location. Such a pattern of effects could suggest that attention during goal-

directed reaching uniquely influences the crossmodal mechanisms underlying the sound-induced flash illusions. In Experiment 2 the influence of directing attention via the eyes and motor system were examined in relation to multisensory perception. Participants fixated on one of three target locations while pointing to the central target. The sensory stimuli could be presented where the participants were aiming and fixating (i.e., visual and action-based attention were aligned), aiming (i.e., action-based attention), fixating (i.e., visual attention), and neither aiming nor fixating. In doing so, it was found that participants were more accurate in judging the number of flashes when either aiming at and fixating to the same location, and when aiming at or fixating to the unique locations where the sensory cues were presented compared with incongruent locations. Critically, the most robust effect for the fusion illusion occurred when aiming and fixation were aligned, further suggesting a role of attention in yielding crossmodal interactions. Overall, the results of the present study may indicate that deploying attention during reaching facilitates multisensory perception at the target relative to nontarget (i.e., unattended) locations.

Experiments 1 and 2 yielded reductions in unimodal and bimodal congruent accuracy across target conditions. Such a finding likely reflects the dedication of processing resources at the target location specifically in relation to engaging in a goal-directed pointing action (see Discussion section of Experiment 1). The dedication of processing resources to the target location (i.e., and the subsequent withdrawal of resources from nontarget locations) had unique effects when considering visual fixation. Typically speaking, overt attention follows the eyes (e.g., Corbetta & Shulman, 2002). When presented within the target of the pointing action, unimodal and bimodal processing was enhanced relative to nontarget locations (i.e., Experiment 1). In Experiment 2, accuracy was greatest when either fixating or directing an action to the location in which the experimental flash/beep stimuli were presented relative to conditions in which no action or fixation was directed. As a result, dividing attention across target locations in relation to the pointing action reduced the availability of attentional resources at nontarget locations. In addition, the common influence of visual fixation and target location for unimodal and bimodal accuracy were consistent with previous work showing that limb movements are coded in an eye-centered reference frame (i.e., Cohen & Andersen, 2000). The results of the reaching paradigm employed were further consistent with the way that attention is directed during goal-directed actions.

In a study conducted by Perry, Sergio, Crawford, and Fallah (2015), activity in area V2 was recorded in nonhuman primates. Area V2 is known to be involved in action-related processes such as reaching, grasping, and attention (for a review, see Perry et al., 2016). Perry et al. (2015) manipulated attention by placing the limb of a nonhuman primate next to

the position where a target would appear (i.e., hand-near condition), or positioned the limb at the side of the animal (i.e., hand-away condition). The target was a rectangle of varied orientation that would be pertinent to hand-related actions such as grasping. Relative to the hand-away condition, the hand-near condition increased cellular response rates in area V2 relating to targets presented in the preferred orientation. Further, the neuronal responses were increased in the preferred orientation and reduced in the orthogonal orientation. Overall, the results of Perry et al. (2015) parallel the previous visual enhancements from oculomotor studies (e.g., Deubel & Schneider, 1996) in that visual processing was enhanced at the locus of attention, in this case the area surrounding the hand. The present study may extend these action-attention related effects to the crossmodal processes underlying the sound-induced flash illusions.

Directing attention during reaching may not influence the susceptibility to the fission illusion, while subsequently reducing the susceptibility to the fusion illusion. Despite differences in both unimodal and bimodal conditions (i.e., the individual auditory and visual components of the fission illusion), fission occurred no differently across target and nontarget conditions. Given the considerable reporting of fission observed here, the fission illusion data supported the hypothesis that top-down processes associated with attention do not contribute significantly to the susceptibility to the fission illusion. Instead, the audiovisual crossmodal processes that integrate the individual illusion components may operate in a bottom-up fashion, independent of attention (i.e., Bertelson et al., 2000; Helbig & Ernst, 2008; Vroomen et al., 2001). However, more work is certainly needed to decipher bottom-up processing from the potential ceiling-level reports of fission reported in the present study.

The results for the fusion stimuli suggest a differing pattern of attentional influence. Susceptibility to the fusion illusion was reduced in the target relative to the nontarget condition, and further facilitated when both aiming and fixating were together at the same spatial location. The influence of target condition revealed in the current study can support the hypothesis that the audiovisual crossmodal processes associated with the fusion illusion occur via the influence of attention, (i.e., top-down processes; see Busse et al., 2005; Donohue et al., 2015; Talsma, 2015; Talsma et al., 2007) and potentially environmental factors such as pointing (see below). The differing pattern of attentional influence found across experiments can provide further support for the hypothesis that fission and fusion arise due to unique, or different, neuropsychological pathways (i.e., Chen et al., 2017). Specifically, bottom-up processes may facilitate the occurrence of fission, and top-down processes may facilitate fusion, in a manner that can be predicted by previous work on attention during goal-directed tasks (e.g., Deubel & Schneider, 1996; Perry et al., 2015).

Top-down and bottom-up contributions to crossmodal perception at the behavioral level is also consistent with previous neurophysiological data. Bolognini et al. (2011) reported that when anodal transcranial direct current stimulation (tDCS) was applied to the temporal cortex, susceptibility to the fission illusion was enhanced, whereas reports of the fusion illusion were inhibited following anodal stimulation of the occipital cortex (i.e., relative to cathodal stimulation). However, anodal and cathodal stimulation had no influence on reports of the fusion illusion. The distinct influence of tDCS on the individual illusions further supports the findings of the current study, as well as others (e.g., Chen et al., 2017), that the sound-induced flash illusions arise due to distinct mechanisms. Speculatively, the fission illusion may arise due to activity within the temporal cortex, while the fusion illusion may be associated with input from the posterior parietal cortex. Indeed, the PPC, which is involved in attention, action, and multisensory perception (e.g., Buneo & Andersen, 2006), is also linked with frontocortical networks associated with dividing attention across space and the senses (see Santangelo, 2018). Neurophysiological evidence should be sought to test the hypothesis that attention influences audiovisual integration in general, and specifically the bimodal combination underlying the fusion illusion.

An additional finding from the present study is in relation to the methodology employed. In previous laboratory investigations, attentional demands were typically altered by employing a dual-task paradigm (i.e., Eg & Behne, 2015; Helbig & Ernst, 2008; Hopkins, Kass, Blalock, & Brill, 2017; Scheidt, Lillis, & Emerson, 2010) or instructing participants to attend to certain modalities or stimulus combinations (i.e., Diaconescu, Alain, & McIntosh, 2011; Mégevand, Molholm, & Nayak, & Foxe, 2013; van der Stoep, Spence, Nijboer, & van der Stigchel, 2015; Taslma, et al., 2007). While such methods have indeed been fruitful for probing the influence of attention on multisensory perception, utilizing an action-based approach to alter attentional demands may provide an additional ecologically based measure to study integration and attention interactions. That is, attention can be deployed to various target or nontarget spatial locations (i.e., distractor locations), as well as objects within the visual scene (i.e., Egly, Driver, & Rafal, 1994; Posner et al., 1980; Tipper et al., 1992). One additional avenue worth pursuing is to further investigate the effect of multisensory distractors on multisensory target processing (e.g., Jensen, Merz, Spence, & Frings, 2019), which can further elucidate multisensory perception mechanisms. Action-based paradigms may be ideally suited to tackle such issues because the direction of effects relating to the influence of attention on multisensory integration can be clearly predicted. That is, directing attention via an action (e.g., Deubel & Schneider, 1996; Tipper et al., 1992) should facilitate information

processes at the attended location. As revealed in the current study, the largest attention-based effects were observed when the target of the reaching action and visual fixation were aligned. As a result, action-based paradigms, may prove useful in future studies aimed at elucidating how attention influences audiovisual integration.

An additional avenue worth pursuing regarding attention and sensory processing relates to the interaction between the two processes and the level at which the interaction occurs. Perhaps the most widely held explanation as to how attention influences multisensory perception is the dedication of cognitive and neural resources to attended stimuli and the associated reduction in the dedication of processing resources for unattended stimuli (e.g., Lavie, 1995; Macaluso et al., 2016). Given the strong coupling of visual fixation and attention (e.g., Corbetta & Schulman, 2002; Treue, 2001), it is further plausible to suggest that attention may impact fundamental visual processing, which would therefore affect multisensory processes. That is, attention could provide a boost to parvocellular activity to encode information in greater detail. Altering parvocellular responses would influence the type of information traveling along the visual pathway, and thus the perception formed by the CNS. Although disentangling the exact mechanisms and neurophysiological structures where attention and multisensory perception interact is beyond the scope of the present manuscript, such behavioral and neurophysiological studies should be pursued in the future.

Overall, the focus of the current study was to investigate how conditions that influence attention—such as eccentricity, cue position, cue probability, reaching position, and trial distribution—yield subsequent effects on multisensory perception. While we are aware that such a pursuit offers many challenges (i.e., Hommel et al., 2019), our results illustrated that deploying attention within an action-based paradigm facilitated unimodal and bimodal perception at the target of the reaching action in comparison to nontarget locations. In addition, spatial attention did not alter the susceptibility to the fission illusion, but yielded greater reports of the fusion illusion at the target relative to the nontarget locations. Such a finding provides further support for the multifaceted interplay between attention and multisensory perception (i.e., Koelewijn, Bronkhorst, & Theeuwes, 2010). In addition, the results of the current study highlight that fission and fusion may arise due to distinct multisensory mechanisms. The influence of attention on each individual illusion pathway provides a consideration for future work examining cognitive influences on multisensory perception.

Acknowledgements Thank you to Joëlle Hajj for collecting some of the data reported in Experiment 2. This study was jointly funded by JST CREST (Grant Number JPMJCR14E4) and JSPS KAKENHI (JP17H00753), the Natural Sciences and Engineering Research Council of Canada, and the University of Toronto Graduate Student Bursary.

Open practices statement Based on the ethics protocols in place, none of the data or materials for the experiments reported here can be made public. Also, none of the experiments was preregistered. However, we would be happy to respond to emails to the corresponding author requesting access to the data reported here.

References

- Andersen, T. S., Tiippana, K., & Sams, M. (2004). Factors influencing audiovisual fission and fusion illusions. *Cognitive Brain Research*, 21(3), 301–308. <https://doi.org/10.1016/j.cogbrainres.2004.06.004>
- Bertelson, P., Vroomen, J., De Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, 62(2), 321–332. <https://doi.org/10.3758/bf03205552>
- Bolognini, N., Rossetti, A., Casati, C., Mancini, F., & Vallar, G. (2011). Neuromodulation of multisensory perception: A tDCS study of the sound-induced flash illusion. *Neuropsychologia*, 49(2), 231–237. <https://doi.org/10.1016/j.neuropsychologia.2010.11.015>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.
- Broadbent, D. E. (1958). Perception and communication. New York: Oxford University Press.
- Buneo, C. A., & Andersen, R. A. (2006). The posterior parietal cortex: Sensorimotor interface for the planning and online control of visually guided movements. *Neuropsychologia*, 44(13), 2594–2606. <https://doi.org/10.1016/j.neuropsychologia.2005.10.011>
- Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences*, 102(51), 18751–18756. <https://doi.org/10.1073/pnas.0507704102>
- Chen, Y. C., Maurer, D., Lewis, T. L., Spence, C., & Shore, D. I. (2017). Central–peripheral differences in audiovisual and visuotactile event perception. *Attention, Perception, & Psychophysics*, 79(8), 2552–2563. <https://doi.org/10.3758/s13414-017-1396-4>
- Cherry, E. C., & Taylor, W. K. (1954). Some further experiments upon the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, 26(4), 554–559. <https://doi.org/10.1121/1.1907373>
- Cohen, Y. E., & Andersen, R. A. (2000). Reaches to sounds encoded in an eye-centered reference frame. *Neuron*, 27(3), 647–652. [https://doi.org/10.1016/S0896-6273\(00\)00073-8](https://doi.org/10.1016/S0896-6273(00)00073-8)
- Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: A time-window-of-integration model. *Journal of Cognitive Neuroscience*, 16(6), 1000–1009. <https://doi.org/10.1162/0898929041502733>
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3), 201–215. <https://doi.org/10.1038/nrn755>
- Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, 36(12), 1827–1837. [https://doi.org/10.1016/0042-6989\(95\)00294-4](https://doi.org/10.1016/0042-6989(95)00294-4)
- Diaconescu, A. O., Alain, C., & McIntosh, A. R. (2011). The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. *Journal of Neurophysiology*, 106(6), 2896–2909. <https://doi.org/10.1152/jn.00303.2011>
- Donohue, S. E., Green, J. J., & Woldorff, M. G. (2015). The effects of attention on the temporal integration of multisensory stimuli. *Frontiers in Integrative Neuroscience*, 9, 1–14. <https://doi.org/10.3389/fnint.2015.00032>
- Downing, C. (1988). Expectancy and visual-spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance*, 14(2), 188–202.
- Eg, R., & Behne, D. M. (2015). Perceived synchrony for realistic and dynamic audiovisual events. *Frontiers in Psychology*, 6, 1–12. <https://doi.org/10.3389/fpsyg.2015.00736>
- Egely, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, 123(2), 161–177. <https://doi.org/10.1037/0096-3445.123.2.161>
- Eriksen, C. W., & Yeh, Y. Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, 11(5), 583–597. <https://doi.org/10.1037/0096-1523.11.5.583>
- Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting Is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4), 1030–1044. <https://doi.org/10.1037/0096-1523.18.4.1030>
- Helbig, H. B., & Ernst, M. O. (2008). Visual-haptic cue weighting is independent of modality-specific attention. *Journal of Vision*, 8(1), 21. <https://doi.org/10.1167/8.1.21>
- Hoffman, J. E., & Nelson, B. (1981). Spatial selectivity in visual search. *Perception & Psychophysics*, 30(3), 283–290. <https://doi.org/10.3758/BF03214284>
- Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J. H., & Welsh, T. N. (2019). No one knows what attention is. *Attention, Perception, & Psychophysics*, 81(7), 2288–2303. <https://doi.org/10.3758/s13414-019-01846-w>
- Hopkins, K., Kass, S. J., Blalock, L. D., & Brill, J. C. (2017). Effectiveness of auditory and tactile crossmodal cues in a dual-task visual and auditory scenario. *Ergonomics*, 60(5), 692–700. <https://doi.org/10.1080/00140139.2016.1198495>
- Huang, L., & Pashler, H. (2005). Attention capacity and task difficulty in visual search. *Cognition*, 94(3), 101–111. <https://doi.org/10.1016/j.cognition.2004.06.006>
- Jensen, A., Merz, S., Spence, C., & Frings, C. (2019). Overt spatial attention modulates multisensory selection. *Journal of Experimental Psychology: Human Perception and Performance*, 45(2), 174–188. <https://doi.org/10.1037/xhp0000595>
- Kahneman, D. (1973). *Attention and effort*. Upper Saddle River, NJ: Prentice Hall.
- Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica*, 134(3), 372–384. <https://doi.org/10.1016/j.actpsy.2010.03.010>
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 21(3), 451–468.
- Loria, T., Hajj, J., Tanaka, K., Watanabe, K., & Tremblay, L. (2018). Visual attention influences audiovisual event perception and the susceptibility to the fusion illusion. *Journal of Exercise Movement and Sport*, 50(1), 46.
- Loria, T., Hajj, J., Tanaka, K., Watanabe, K., & Tremblay, L. (2019). The deployment of spatial attention during goal-directed action alters audio-visual integration. *Journal of Vision*, 19(10), 111.
- Macaluso, E., Noppeney, U., Talsma, D., Vercillo, T., Hartcher-O'Brien, J., & Adam, R. (2016). The curious incident of attention in multisensory integration: Bottom-up vs. top-down. *Multisensory Research*, 29(6), 557–583. <https://doi.org/10.1163/22134808-00002528>
- Mégevand, P., Molholm, S., Nayak, A., & Foxe, J. J. (2013). Recalibration of the multisensory temporal window of integration results from changing task demands. *PLOS ONE*, 8(8), e71608. <https://doi.org/10.1371/journal.pone.0071608>
- Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a

- sound-induced visual illusion. *Journal of Neuroscience*, 27(15), 4120–4131. <https://doi.org/10.1523/JNEUROSCI.4912-06.2007>
- Mishra, J., Martinez, A., & Hillyard, S. A. (2010). Effect of attention on early cortical processes associated with the sound-induced extra flash illusion. *Journal of Cognitive Neuroscience*, 22(8), 1714–1729. <https://doi.org/10.1162/jocn.2009.21295>
- Müller, H. J., & Findlay, J. M. (1987). Sensitivity and criterion effects in the spatial cueing of visual attention. *Attention, Perception, & Psychophysics*, 42(4), 383–399. <https://doi.org/10.3758/BF03203097>
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1):97–113
- Osterberg, G. (1935). Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica* 6, 1–103.
- Palmer, B., Nasman, V. T., & Wilson, G. F. (1994). Task decision difficulty: Effects on ERPs in a same-different letter classification task. *Biological Psychology*, 38(2/3), 199–214. [https://doi.org/10.1016/0301-0511\(94\)90039-6](https://doi.org/10.1016/0301-0511(94)90039-6)
- Perry, C. J., Sergio, L. E., Crawford, J. D., & Fallah, M. (2015). Hand placement near the visual stimulus improves orientation selectivity in V2 neurons. *Journal of Neurophysiology*, 113(7), 2859–2870. <https://doi.org/10.1152/jn.00919.2013>
- Perry, C. J., Amarasekariya, P., & Fallah, M. (2016). An eye in the palm of your hand: Alterations in visual processing near the hand, a mini-review. *Frontiers in Computational Neuroscience*, 10, 1–8. <https://doi.org/10.3389/fncom.2016.00037>
- Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109(2), 160–174. <https://doi.org/10.1037/0096-3445.109.2.160>
- Santangelo, V. (2018). Large-scale brain networks supporting divided attention across spatial locations and sensory modalities. *Frontiers in Integrative Neuroscience*, 12. <https://doi.org/10.3389/fnint.2018.00008>
- Scheidt, R. A., Lillis, K. P., & Emerson, S. J. (2010). Visual, motor and attentional influences on proprioceptive contributions to perception of hand path rectilinearity during reaching. *Experimental Brain Research*, 204(2), 239–254. <https://doi.org/10.1007/s00221-010-2308-1>
- Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews*, 7(3), 269–284. <https://doi.org/10.1016/j.plrev.2010.04.006>
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, 408(6814), 788. <https://doi.org/10.1038/35048669>
- Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, 59(1), 1–22. <https://doi.org/10.3758/BF03206843>
- Talsma, D. (2015). Predictive coding and multisensory integration: An attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience*, 9, 1–13. <https://doi.org/10.3389/fnint.2015.00019>
- Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*, 17(3), 679–690. <https://doi.org/10.1093/cercor/bhk016>
- Tipper, S. P., Lortie, C., & Baylis, G. C. (1992). Selective reaching: Evidence for action-centered attention. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4), 891–905. <https://doi.org/10.1037/0096-1523.18.4.891>
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., & Théoret, H. (2007). Speech and non-speech audio-visual illusions: A developmental study. *PLOS ONE*, 2(8). <https://doi.org/10.1371/journal.pone.0000742>
- Tremblay, L., & Nguyen, T. (2010). Real-time decreased sensitivity to an audio-visual illusion during goal-directed reaching. *PLOS ONE*, 5(1), 2–5. <https://doi.org/10.1371/journal.pone.0008952>
- Treue, S. (2001). Neural correlates of attention in primate visual cortex. *Trends in Neurosciences*, 24(5), 295–300. [https://doi.org/10.1016/S0030-6657\(08\)70226-9](https://doi.org/10.1016/S0030-6657(08)70226-9)
- van Atteveldt, N. M., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: Flexible use of general operations. *Neuron*, 81(6), 1240–1253. <https://doi.org/10.1016/j.neuron.2014.02.044>
- van der Stoep, N., Spence, C., Nijboer, T. C. W., & 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. <https://doi.org/10.1016/j.actpsy.2015.09.010>
- Vroomen, J., Bertelson, P., & de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Attention, Perception, & Psychophysics*, 63(4), 651–659.

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