Psychon Bull Rev (2011) 18:904–909
DOI 10.3758/s13423-011-0131-8

BRIEF REPORT

Multisensory enhancement of attentional capture in visual search

Pawel J. Matusz · Martin Eimer

Published online: 12 July 2011

Abstract Multisensory integration increases the salience of sensory events and, therefore, possibly also their ability to capture attention in visual search. This was investigated in two experiments where spatially uninformative color change cues preceded visual search arrays with color-defined targets. Tones were presented synchronously with these cues on half of all trials. Spatial-cuing effects indicative of cue-triggered capture of attention were larger on tone-present than on tone-absent trials, demonstrating multisensory enhancements of attentional capture. Larger capture effects for audiovisual events were found when cues were color singletons, and also when they appeared among heterogeneous color distractors. Tone-induced increases of attentional capture were independent of color-specific top-down task sets, suggesting that this multisensory effect is a stimulus-driven bottom-up phenomenon.

Keywords Attentional capture · Multisensory integration · Bottom-up salience · Visual search

Spatial attention is critically involved in the detection, selection, and identification of targets in multistimulus visual scenes. When stimuli compete for visual processing, attention acts to resolve this competition either in a bottom-up way in favor of the most salient item or in a top-down fashion in favor of currently task-relevant events (Desimone & Duncan, 1995). The relative roles of bottom-up and top-down factors in visual search remain contentious: Some have argued that highly salient but task-irrelevant visual stimuli will capture attention in an automatic bottom-up fashion even when another task-relevant stimulus is simultaneously present (e.g., Theeuwes, 1991, 2010), while others believe that attentional capture by salient visual events is always under the control of top-down task sets (e.g., Folk, Remington, & Johnson, 1992).

In the visual domain, salience is usually defined in terms of strong local contrast signals, which can be triggered by the presence of a feature singleton (e.g., the only red stimulus among gray distractors in a visual search array). However, increased salience can also be a product of multisensory integration. Evidence that multisensory interactions boost neural processing and increase perceptual salience has come from a number of sources. Enhanced neural responses to synchronous multisensory, as compared with unimodal, stimulation have been observed in different brain regions, including the superior colliculus (e.g., Meredith, Nemitz, & Stein, 1987). ERP studies (e.g., Giard & Perronet, 1999) have provided evidence for the rapid onset of audiovisual enhancements of cortical processing. In line with such physiological results, behavioral studies have shown that concurrent audiovisual stimulation increases perceived visual brightness (Stein, London, Wilkinson, & Price, 1996) and decreases visual contrast thresholds (Lippert, Logothetis, & Kayser, 2007; see also Gillmeister & Eimer, 2007, for analogous audiotactile effects), suggesting that multisensory integration also enhances perceptual salience.

By generating a salience-driven selection bias in favor of visual stimuli that are accompanied by an event in another sensory modality, multisensory integration might affect attentional competition in vision. Until now, the evidence in favor of attentional benefits for multisensory events has been mixed. Experiments using uninformative exogenous spatial cues have shown that differences in the magnitude

P. J. Matusz · M. Eimer (✉)
Department of Psychological Sciences, Birkbeck College, University of London,
Malet Street,
London WC1E 7HX, UK
e-mail: m.eimer@bbk.ac.uk

of attentional capture triggered by visual, auditory, and audiovisual cues may depend on perceptual load (Santangelo & Spence, 2007). Vroomen and De Gelder (2000) studied serial presentations of audiovisual events and observed an increase in the subjective duration of visual targets when they appeared concurrently with a deviant tone (a frequency singleton). This *freezing effect* was associated with improved target detection performance. Because single-stream serial presentation paradigms cannot provide insights into the impact of multisensory integration on attentional competition in multistimulus arrays, Van der Burg, Olivers, Bronkhorst, and Theeuwes (2008) used a modified visual search paradigm to study this issue. Participants searched for a vertical or horizontal target line among numerous oblique distractors. Some search array items changed color at random intervals. Critically, a target color change was accompanied by a tone in some blocks, but not in others. Target detection was substantially faster in tone-present blocks, relative to tone-absent ones, in line with multisensory facilitation of attentional target selection. Van der Burg et al. suggested that visual salience is boosted through audiovisual integration, resulting in automatic attentional capture by audiovisual events (*pip and pop* effect). However, it remained unclear whether this effect represents a genuine bottom-up phenomenon. Reaction times (RTs) were much longer to targets in tone-present blocks than in typical pop-out visual search tasks, and search slopes were not flat, suggesting that search was difficult and serial. Furthermore, the effectiveness of sounds in improving search efficiency was modulated by the likelihood of their co-occurrence with target or distractor events (Van der Burg et al., 2008, Experiment 4), which points to an important role for top-down control.

The aim of the present study was to provide more direct evidence for a link between multisensory integration and selective attention under conditions where attentional capture by audiovisual events is rapid and involuntary. We adapted the spatial-cuing procedure introduced by Folk et al. (1992) for a multisensory context. On each trial, a color change cue preceded a visual search display that contained a color-defined target. Critically, these cues were accompanied by a tone on 50% of all trials. Simultaneously with tone onset, one item in the cue array changed color, relative to the immediately preceding base array (Fig. 1). Color change location was nonpredictive with respect to the location of subsequent targets. Spatial-cuing effects (i.e., shorter RTs to visual search targets at cued, relative to uncued, locations) were measured as an index of rapid involuntary attentional capture by these cues. The central question was whether such spatial-cuing effects would be boosted in tone-present trials, relative to tone-absent ones, as would be expected if multisensory integration enhances the salience of visual events and, thus, their ability to capture attention.

To assess the role of top-down task set for multisensory enhancements of attentional capture, search task requirements were varied between experiments. In Experiment 1, search arrays contained a color singleton target bar among gray distractors. Two different color targets were used, which appeared in random order and with equal probability. The color change cue could match one of the two target colors or could be a change to a nontarget color. A previous study (Eimer & Kiss, 2010) demonstrated that when the color of target singletons is unpredictable, participants adopt a "singleton search mode" (Bacon & Egeth, 1994), and color singleton cues capture attention irrespective of their feature value, resulting in spatial-cuing effects for target color as well as nontarget color cues. If multisensory integration boosts attentional capture, larger spatial-cuing effects should be found in tone-present trials, relative to tone-absent trials, for both types of cue. In Experiment 2, bars in one predefined target color were presented among five differently colored distractor bars. Because targets were no longer feature singletons, participants were now forced to adopt a color-specific "feature search mode" (Bacon & Egeth, 1994; Eimer, Kiss, Press, & Sauter, 2009; Lamy, Leber, & Egeth, 2004). Target color or nontarget color change cues were presented against a background of heterogeneous color distractors (Fig. 1, bottom). With a feature-specific top-down task set, spatial-cuing effects indicative of attentional capture should now be triggered by target color cues only (Eimer et al., 2009; Lamy et al., 2004). One question was whether a boost of capture by multisensory integration on tone-present trials would still be observed for heterogeneous (and thus less salient) color change cue arrays. The other question was whether any audiovisual enhancement of capture would depend on color-specific top-down task sets. If this effect is a genuine bottom-up phenomenon, it should be observed for all color change cues. If it is contingent on top-down task set, it should be elicited only by target color cues.

Method

Participants Twenty paid volunteers (19–31 years of age, mean age 24.7 years; 9 males) took part in Experiment 1. Twenty-five volunteers (19–40 years of age, mean age 27.5 years; 11 males) participated in Experiment 2. One participant in Experiment 2 was excluded due to an inability to perform the task as instructed, and two others because their mean RTs were more than 2 *SD*s longer than the group mean. All participants had normal or corrected-to-normal vision, and all but 5 were right-handed.

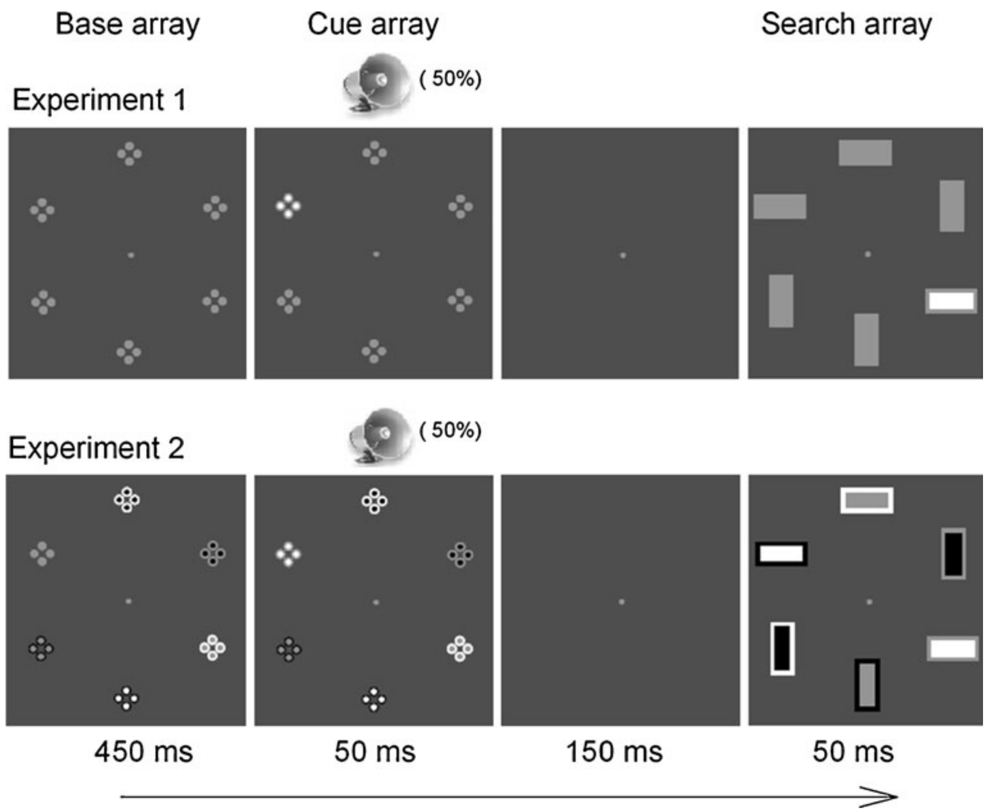

Fig. 1 Stimulus setup and trial sequence used in Experiment 1 (top) and Experiment 2 (bottom). Base arrays were followed by cue arrays that included a color change and were accompanied by a synchronous tone on 50% of all trials. Search arrays with a color-defined target followed after a 150-ms interval. In Experiment 1, color change cues and targets were color singletons. In Experiment 2, they appeared among heterogeneous color distractors. Different colors are indicated by different combinations of bar and outline shadings

Stimuli, procedure, and analysis Experiments were conducted in a dimly lit, air-conditioned room, with participants seated at a distance of 100 cm from a 17-in. CRT monitor (75-Hz refresh rate). On each trial, a search array containing a target was preceded by a color change cue (one item in the base array changing color in the cue array). This cue could be accompanied by a simultaneous tone (Fig. 1). Base arrays, cue arrays, and search arrays all contained six elements that were presented equidistantly along the circumference of an imaginary circle against a black background, at an angular distance of 2.1° from a central fixation point. All visual stimuli were approximately equiluminant (~10.5 cd/m^2).

In Experiment 1, search arrays contained five gray bars and one color singleton target bar, each subtending 0.7° × 0.3°. For each participant, the target bar was equally likely to have one of two colors (green and blue, green and red, or red and blue), with target color set counterbalanced across participants. CIE *x*/*y* chromaticity coordinates were .285/.591, .161/.128, and .621/.343 for green, blue, and red targets, respectively. In Experiment 2, search displays contained distractor bars in five different task-irrelevant colors plus one target color bar. Distractor bar color was randomly chosen from a set of six colors (CIE chromaticity coordinates: purple, .220/.119; turquoise, .248/.429; green, .285/.591; pink, .493/.281; orange, .558/ .387; yellow, .432/.485), with the constraint that each bar was assigned a different color. For half of all participants, the target color was red; for the other half, it was blue. In both experiments, the orientation of each bar in the search array was randomly determined on each trial. Color-defined target bars were presented with equal probability and in a random order at one of the four lateral locations of the search array, but never at the top or bottom location. Participants' task was to detect the target bar and to respond to its orientation (horizontal or vertical) by pressing one of two vertically oriented response keys with their left or right index finger (upper key for vertical targets, lower key for horizontal targets). Key–hand assignment was reversed after half of all blocks. Participants were instructed to respond as quickly and accurately as possible.

To generate color change cues that could be synchronized with tones, base arrays (450-ms duration) were presented at the beginning of each trial and were immediately replaced by cue arrays (50-ms duration). In Experiment 1, each base array consisted of six sets of four closely aligned gray dots (subtending 0.1° × 0.1° of visual angle; CIE chromaticity coordinates, .308/.345). In the cue array, one of the four lateral sets was equiprobably blue, green, or red, resulting in a color change. For each participant, two of the cue colors matched the two possible target colors (target color cues), while the third color did not (nontarget color cues). In Experiment 2, the color change cue appeared against a background of heterogeneously colored items, in order to equate the color features of cue and search arrays. Each item in the base array was randomly assigned one of the six colors that were used for distractors in the search array. Cue arrays were identical to base arrays, except that one of the items on the left or right side changed color. The color change in the cue display could be either the appearance of the target color or the appearance of a nontarget color that did not appear in the search array. For participants searching for blue target bars, target color cues were blue and nontarget color cues red, and vice versa for participants searching for red targets. In both experiments, target color and nontarget color cues appeared in random order and with equal probability and were spatially uninformative with respect to the location of a subsequent color-defined target bar.

On half of all trials, the color change in the cue array coincided with the onset of a pure sine-wave tone (2000-Hz frequency) that was presented for 50 ms from a loudspeaker located at the top of the monitor. In Experiment 1, tones had an intensity of 65 dB SPL, as measured from a position adjacent to participants' ears. In Experiment 2, tone intensity was 80 dB[1]. Tone-present and tone-absent trials were randomly intermixed. The interval between cue array offset and search array onset was 150 ms, and the intertrial interval was 1,500 ms. Experiment 1 contained 12 blocks of 48 trials, resulting in a total of 576 trials. Experiment 2 contained 8 blocks with 64 trials each, resulting in a total of 512 trials. Participants completed 1 training block at the start of the experiment and 1 after half of all blocks, when the new key–hand mapping was introduced. They were informed that tones and color change cues were spatially uninformative and task irrelevant and were instructed to ignore them.

RTs and error rates were analyzed in three-way ANOVAs with cue color (target color cue vs. nontarget color cue), tone presence (tone-present vs. tone-absent trials), and spatial cuing (target at cued location vs. one of the three uncued locations) as within-subjects factors.

Results

Experiment 1

Trials with RTs below 200 ms and above 1,000 ms were excluded from analyses (1% of all trials). Figure 2 shows RTs (line graphs) and error rates (bar graphs) for targets at cued and uncued locations, separately for tone-present and tone-absent trials and for trials with target color or nontarget color cues.

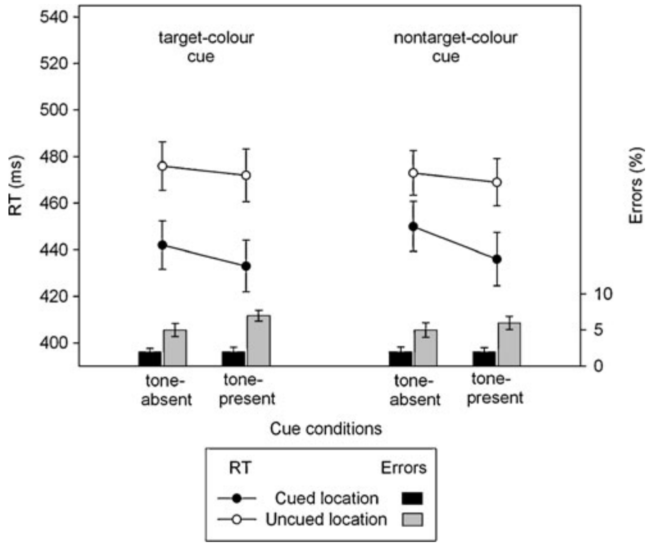

Fig. 2 Mean reaction times (RTs; line graphs) and error rates (bar graphs) in Experiment 1 in response to targets at cued and uncued locations, shown separately for trials with target color cues (left) and nontarget color cues (right) and for tone-present and tone-absent trials. Error bars represent standard errors of the means

RTs on tone-present trials were shorter than RTs on tone-absent trials (452 vs. 460 ms), resulting in a main effect of tone presence, $F(1, 19) = 13.84$, $p < .001$, $\eta_p^2 = .42$. RTs were shorter for targets at cued, as compared with uncued, locations (440 vs. 472 ms), as reflected by a main effect of spatial cuing, $F(1, 19) = 172.86$, $p < .001$, $\eta_p^2 = .91$. No cue color × spatial cuing interaction was present, since cuing effects of similar size were triggered by target color and nontarget color cues (see Fig. 2). Most important, there was an interaction between tone presence and spatial cuing, $F(1, 19) = 4.71$, $p < .05$, $\eta_p^2 = .2$. Spatial-cuing effects were larger on tone-present (36 ms) than on tone-absent (29 ms) trials. This audiovisual enhancement of attentional capture was unaffected by cue color (target color cues versus nontarget color cues), $F < 1$.

Errors were more frequent on tone-present than on tone-absent trials (4.2% vs. 3.4%), resulting in a main effect of tone presence, $F(1, 19) = 5.49$, $p < .05$, $\eta_p^2 = .22$. There was a main effect of spatial cuing, $F(1, 19) = 57.62$, $p < .001$, $\eta_p^2 = .75$, since more errors occurred in response to targets at uncued locations, relative to cued targets. A tone presence × spatial cuing interaction, $F(1, 19) = 4.79$, $p < .05$, $\eta_p^2 = .2$, was due to the fact that the increase in error rates for uncued, as compared with cued, targets was more pronounced for tone-present trials (Fig. 2). There were no other significant main effects or interactions for error rates.

[1] Tone intensity was increased, relative to Experiment 1, because a pilot study showed no reliable multisensory enhancement of attentional capture with heterogeneous color change cue arrays and 65-dB tones. Color change cues that appear in the context of heterogeneous color distractors are less salient than the singleton color change cues used in Experiment 1, and more intense concurrent tones may therefore be required to trigger an audiovisual boost of capture.

Experiment 2

Trials with RTs below 200 ms and above 1,000 ms were excluded from analyses (fewer than 1% of all trials). Figure 3 shows RTs (line graphs) and error rates (bar graphs) for targets at cued and uncued locations, separately for tone-present and tone-absent trials and for trials with target color or nontarget color cues.

As in Experiment 1, RTs were shorter on tone-present trials, relative to tone-absent ones (480 vs. 495 ms), resulting in a main effect of tone presence, $F(1, 21) = 32.83$, $p < .001$, $\eta_p^2 = .61$. A main effect of spatial cuing, $F(1, 21) = 47.49$, $p < .001$, $\eta_p^2 = .69$, reflected shorter RTs to targets at cued, as compared with uncued, locations (479 vs. 496 ms). In contrast to Experiment 1, there was now a significant cue color × spatial cuing interaction, $F(1, 21) = 39.37$, $p < .001$, $\eta_p^2 = .65$, indicative of task-set-contingent attentional capture. Target color cues triggered spatial-cuing effects, $F(1, 21) = 55.2$, $p < .001$, whereas no such effects were elicited by nontarget color cues (Fig. 3). Most important, there was again a reliable interaction between tone presence and spatial cuing, $F(1, 21) = 4.5$, $p < .05$, $\eta_p^2 = .18$, with larger cuing effects on tone-present trials, relative to tone-absent ones (20 vs. 11 ms, collapsed across target color and nontarget color cues). There was no evidence for any cue color × tone presence × spatial cueing interaction, $F < 1$, indicating that the audiovisual enhancement of attentional capture was independent of color-specific top-down task sets.

As in Experiment 1, incorrect responses were more frequent on tone-present than on tone-absent trials (5.7% vs. 4.3%), $F(1, 21) = 10.65$, $p < .001$, $\eta_p^2 = .34$, and more frequent for targets at uncued locations, relative to targets at cued locations (6.1% vs. 3.8%), $F(1, 21) = 21.23$, $p < .001$, $\eta_p^2 = .51$. There was an interaction between cue color and spatial cuing, $F(1, 21) = 8.97$, $p < .01$, $\eta_p^2 = .29$, since the increase in error rate for uncued target locations, relative to cued target locations, was larger for trials with target color cues than for trials with nontarget color cues. There were no other significant main effects or interactions for error rates.

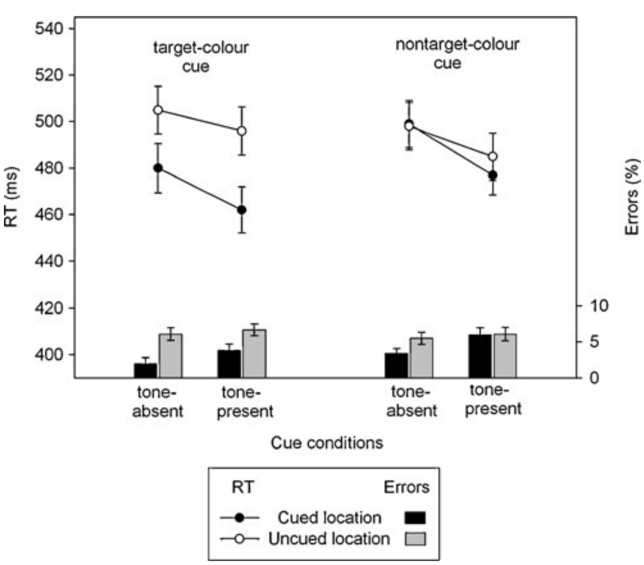

Fig. 3 Mean reaction times (RTs; line graphs) and error rates (bar graphs) in Experiment 2 in response to targets at cued and uncued locations, shown separately for trials with target color cues (left) and nontarget color cues (right) and for tone-present and tone-absent trials. Error bars represent standard errors of the means

Discussion

The present results demonstrate that multisensory integration can boost attentional capture by task-irrelevant sensory events. Spatially uninformative color change cues triggered spatial-cuing effects on RTs to subsequent color-defined targets, indicative of involuntary attentional capture. Critically, these spatial-cuing effects were reliably larger on trials on which these cues were accompanied by a simultaneous tone than on tone-absent trials. This was the case not only when cue presentation involved the sudden appearance of a unique color singleton among gray distractors (Experiment 1), but also for color changes that took place against a background of heterogeneously colored distractors (Experiment 2). Multisensory integration can enhance the salience of sensory events (e.g., Stein et al., 1996), and the increase in the magnitude of spatial-cuing effects for audiovisual, as compared with visual, cues is likely to be a direct reflection of this fact: Multisensory events capture attention more strongly because they are more salient than unimodal visual events. Previous studies of multisensory effects on attention have investigated single-stream RSVP displays (Vroomen & de Gelder, 2000) or difficult search tasks with long RTs (Van der Burg et al., 2008). The present study demonstrated that concurrent tones can enhance the capacity of visual events to capture attention in multistimulus search arrays under conditions in which capture is rapid and involuntary. These results therefore provide direct evidence that multisensory integration can have an immediate impact on attentional selectivity in visual search.

In both experiments, target RTs were shorter and errors more frequent on tone-present than on tone-absent trials, indicating that tones produced alerting effects, reflected by a speed–accuracy trade-off for target detection and identification. One could argue that tone-induced alerting may also have enhanced attentional orienting (see Callejas, Lupiáñez, & Tudela, 2004, for larger spatial-cuing effects after alerting signals) and that it was this fact, rather than multisensory integration, that was responsible for increased attentional capture effects on trials with audiovisual cues. To test this alternative interpretation, a follow-up experiment with 12 participants was conducted. This experiment was identical to Experiment 1, except that tones and visual cues were now separated by a 450-ms interval, because tones appeared synchronously with base array onset. Under these conditions, multisensory integration will be eliminated, but auditory alerting effects should remain. As was expected, a reliable spatial-cuing effect was obtained, $F(1, 11) = 31.0$, $p < .001$. Critically, this effect was not modulated by tone presence, $F < 1$, and cuing effects on tone-present trials (25 ms) were, in fact, numerically smaller than cuing effects on tone-absent trials (27 ms). This result rules out a critical role of tone-induced alerting for audiovisual enhancements of attentional capture.

If multisensory integration increases bottom-up stimulus salience, and if such an increase in salience is responsible for stronger attentional capture by color change cues on tone-present trials, one would expect this multisensory effect to be unaffected by top-down task sets. In both experiments, audiovisual enhancements of attentional capture were indeed independent of cue color. In Experiment 1, where participants searched for one of two equiprobable color singleton targets, spatial-cuing effects indicative of attentional capture were triggered by target color as well as nontarget color cues, suggesting a color-unspecific singleton search mode (Bacon & Egeth, 1994; Eimer & Kiss, 2010). In Experiment 2, where targets were no longer feature singletons, spatial-cuing effects were restricted to target color cues, demonstrating that participants now adopted a color-specific feature search mode. In spite of this fact, the multisensory enhancement of attentional capture was not modulated by cue color. This observation provides initial evidence for the hypothesis that this effect represents a primarily salience-driven bottom-up phenomenon that may be triggered largely irrespective of top-down search goals.

Finally, it should be noted that even though audiovisual enhancements of capture were reliable, these effects were considerably smaller than the generic capture effects observed in both experiments, suggesting that multisensory boosts of stimulus salience are less important for attentional capture in visual search than are local feature differences and top-down task sets. It will be important to find out whether multisensory enhancements of attentional capture are larger under conditions where events from different sensory modalities are not just temporally, but also spatially aligned.

References

Bacon, W. F., & Egeth, H. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics, 55,* 485–496.

Callejas, A., Lupiáñez, J., & Tudela, P. (2004). The three attentional networks: On their independence and interactions. *Brain and Cognition, 54,* 225–227.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.

Eimer, M., & Kiss, M. (2010). Top-down search strategies determine attentional capture in visual search: Behavioural and electrophysiological evidence. *Attention, Perception, & Psychophysics, 72,* 951–962.

Eimer, M., Kiss, M., Press, C., & Sauter, D. (2009). The roles of feature-specific task set and bottom-up salience in attentional capture: An ERP study. *Journal of Experimental Psychology. Human Perception and Performance, 35,* 1316–1328.

Folk, C. L., Remington, R., & Johnson, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology. Human Perception and Performance, 19,* 676–681.

Giard, M. H., & Perronet, F. (1999). Auditory–visual integration during multimodal object recognition in humans: A behavioural and electrophysiological study. *Journal of Cognitive Neuroscience, 11,* 473–490.

Gillmeister, H., & Eimer, M. (2007). Tactile enhancement of auditory detection and perceived loudness. *Brain Research, 1160,* 58–68.

Lamy, D., Leber, A. B., & Egeth, H. E. (2004). Effects of stimulus-driven salience within feature search mode. *Journal of Experimental Psychology. Human Perception and Performance, 30,* 1019–1031.

Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research, 1173,* 102–109.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons: I. Temporal factors. *Journal of Neuroscience, 7,* 3215–3229.

Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology. Human Perception and Performance, 33,* 1311–1321.

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,* 497–506.

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics, 50,* 184–193.

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica, 135,* 77–99.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Non-spatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1053–1065.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organisation on vision. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1583–1590.