Psychophysiology, 50 (2013), 996–1009. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/psyp.12086

Top-down control of audiovisual search by bimodal search templates

PAWEL J. MATUSZ AND MARTIN EIMER
Department of Psychological Sciences, Birkbeck College, University of London, London, UK

Abstract

To test whether the attentional selection of targets defined by a combination of visual and auditory features is guided in a modality-specific fashion or by control processes that are integrated across modalities, we measured attentional capture by visual stimuli during unimodal visual and audiovisual search. Search arrays were preceded by spatially uninformative visual singleton cues that matched the current target-defining visual feature. Participants searched for targets defined by a visual feature, or by a combination of visual and auditory features (e.g., red targets accompanied by high-pitch tones). Spatial cueing effects indicative of attentional capture were reduced during audiovisual search, and cue-triggered N2pc components were attenuated and delayed. This reduction of cue-induced attentional capture effects during audiovisual search provides new evidence for the multimodal control of selective attention.

Descriptors: Visual attention, Top-down control, Crossmodal attention, Multisensory integration, Event-related potentials

Only a small subset of the information that is entering our senses at each moment can be fully processed. Because external objects and events constantly compete with each other for access to perception and action control, effective behavioral control is critically dependent on attentional mechanisms that bias this competition in favor of objects that are important to behavioral goals (Desimone & Duncan, 1995). The selection of these objects is controlled by working memory representations or "attentional templates" of currently goal-relevant features of the environment (e.g., Carlisle, Arita, Pardo, & Woodman, 2011; Duncan & Humphreys, 1989; Olivers & Eimer, 2011). Most investigations into top-down attentional control by attentional templates have focused on the visual modality and on tasks where task-relevant stimuli are defined in terms of one specific feature or feature dimension (e.g., "red" or "any color discontinuity"; e.g., Bacon & Egeth, 1994; Eimer & Kiss, 2008; Eimer, Kiss, Press, & Sauter, 2009; Folk, Remington, & Johnston, 1992; Lamy, Leber, & Egeth, 2004). For example, Folk et al. (1992) used spatial cueing procedures to demonstrate that salient visual singletons capture attention only when they match a currently active task set, but not when their features are task irrelevant (task-set contingent attentional capture; see also Folk & Remington, 1998). However, attentional selectivity in naturalistic environments is rarely directed towards single elementary perceptual features (e.g., "red" or "round"). In the real world, we typically search for objects that are defined by a conjunction of features from different dimensions (e.g., search for a black, small, and rectangular mobile phone). Importantly, we often also use simultaneous cues from different sensory modalities to locate targets. When we want to find our misplaced mobile phone while it is ringing, attentional templates will include both visual (color, size, shape) as well as auditory features (the pitch or melody of the ringtone).

The question how such multifeature attentional templates are organized has so far rarely been addressed. One possibility is that each target-defining feature is represented separately and independently. Alternatively, attentional templates might represent different features of currently task-relevant objects as a single integrated object representation. Most current models of visual attention (Treisman & Gelade, 1980; Wolfe, 1994, 2007) assume that visual search is guided independently by separate representations of task-relevant features. According to the Guided Search model (Wolfe, 1994, 2007), the allocation of attention is controlled by a central spatiotopically organized salience map that receives inputs from anatomically separate and independently operating visual feature channels. Top-down attentional control is implemented through the task-dependent weighting of these inputs, with weights being applied independently to each channel, and each channel then contributing independently and additively to the overall activation profile of the salience map. The alternative hypothesis that attention is guided by fully integrated representations of target objects is consistent with experimental evidence for object-based selection (e.g., Duncan, 1984), and is also in line with evidence that visual working memory (where attentional templates are maintained) represents integrated objects rather than individual features of objects (Luck & Vogel, 1997).

Attentional templates guide search not just for visual targets, but may also be involved in search for targets that are defined

This research was funded by a Birkbeck College Research Studentship awarded to PJM. The authors thank Monika Kiss for comments on earlier versions of this manuscript.

Address correspondence to: Martin Eimer, Department of Psychological Sciences, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK. E-mail: m.eimer@bbk.ac.uk

PSYCHOPHYSIOLOGY

across different sensory modalities. In real-world environments, search is often directed towards audiovisual target objects (e.g., the mobile phone with its personalized ringtone). In such situations, a plausible assumption is that attention is controlled by anatomically and functionally independent visual and auditory representations of target-defining features, which affect processing in modality-specific brain regions: Visual target features attract visual attention and produce spatially selective modulations in visual brain areas, while simultaneously present auditory target features attract auditory attention and affect processing in auditory cortex. If this were correct, the attentional processing of visual information in visual cortex should be unaffected by the simultaneous presence versus absence of auditory target features. An alternative possibility is that the attentional mechanisms that are involved in search for audiovisually defined targets do not operate in a strictly modality-specific fashion, but are instead linked. In this case, the ability of visual target stimuli to attract attention should be influenced by whether an auditory task-relevant stimulus is also present. In the extreme case, audiovisual search might be guided by an attentional template where target-defining visual and auditory features are fully integrated. If this were the case, task-set matching visual objects would capture attention only when they are accompanied by the relevant auditory target feature, but not when they are presented in isolation. Alternatively, visual and auditory features may be partially integrated but retain some degree of independence. In this case, attentional capture would not be completely eliminated but still reduced during audiovisual search when visual target objects appear in the absence of task-set matching auditory stimuli.

The question whether the guidance of attention during search for crossmodally defined targets is organized in a strictly modality-specific fashion or is linked across sensory modalities is not addressed by current models of visual search, which are exclusively concerned with the role of visual features and dimensions in the control of attentional selectivity. This question has so far also been ignored by studies of crossmodal attention and multisensory integration, which have typically focused on spatial synergies of attention across sensory modalities (cf. Eimer, van Velzen, & Driver, 2002; Spence & Driver, 1996), the automatic capture of attention by synchronous crossmodal events (e.g., Matusz & Eimer, 2011; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008), or the preferential selection of naturalistic visual objects that are accompanied by their characteristic sound (e.g., Iordanescu, Grabowecky, Franconeri, Theeuwes, & Suzuki, 2010). In the present study, we investigated whether the attentional selection of visual stimuli during audiovisual search where targets are defined by a combination of visual and auditory features differs from attentional selection processes that are elicited during purely unimodal visual search.

Task-set contingent attentional capture effects can provide direct insights into the processes that guide attention during visual and audiovisual search. In the original unimodal visual version of the contingent capture paradigm (Folk et al., 1992), spatially non-informative feature singleton cues preceded search arrays with feature-defined targets. Responses to targets were faster when targets were presented at the same location as the preceding cue, and these spatial cueing effects were interpreted as evidence that the cues had captured attention. Critically, such attentional capture effects were observed only under conditions where the cues matched currently task-relevant features, but not in tasks where the perceptual properties of the cues were task irrelevant. This pattern of results demonstrates that the capacity of singleton cues to attract attention is contingent on top-down task sets. In the present study, we investigated whether visual feature singleton cues fully retain their ability to capture attention in a task-set contingent fashion when targets are defined by a combination of visual and auditory features. The singleton cues always matched the currently task-relevant visual feature, and should therefore attract attention during unimodal search (e.g., red singleton cues should capture attention during search for red targets). The critical question was whether this would still be the case when targets were defined by a combination of visual and auditory features. For example, do red singleton cues capture attention during search for red bars that are accompanied by high-pitch tones? If top-down control of attention operated in a strictly modality-specific fashion, cue-induced attentional capture effects should be identical during unimodal and audiovisual tasks. In contrast, if there are crossmodal links in the guidance of selective attention that affect the attentional processing of visual events, the ability of visual cues to attract attention should be reduced (or even entirely eliminated) during audiovisual search, because these cues only partially match the features of audiovisually defined target objects.

To assess the capacity of color singleton cues to capture attention in unimodal visual and audiovisual task contexts, we measured reaction times (RTs) to subsequent targets, and recorded event-related brain potentials (ERPs) during task performance. Behaviorally, cue-induced attentional capture was assessed by spatial cueing effects (i.e., faster RTs in response to targets that appear at the same location as the cue relative to targets at other uncued locations). To obtain an online electrophysiological marker of attentional capture by color singleton cues, we measured the N2pc component in response to these cues. The N2pc is an enhanced negativity that emerges around 200 ms after the onset of visual arrays with a candidate target stimulus over posterior scalp electrodes contralateral to the side of this stimulus, and is linked to the spatial selection of potential targets among distractors in visual search tasks (Eimer, 1996; Luck & Hillyard, 1994; Mazza, Turatto, Umiltà, & Eimer, 2007; Woodman, Arita, & Luck, 2009). Recent studies have shown that the N2pc can track rapid attentional capture by visual feature singletons (e.g., Hickey, McDonald, & Theeuwes, 2006), and demonstrate the task-set contingent nature of this capture (e.g., Eimer & Kiss, 2008; Eimer et al., 2009; Kiss, Grubert, Petersen, & Eimer, 2012; Leblanc, Prime, & Jolicoeur, 2008; Lien, Ruthruff, Goodin, & Remington, 2008).

Experiment 1 included three task conditions (see Figure 1). In all tasks, a cue array that contained one colored item among five gray background items was followed by a search array that contained a color singleton bar among five other gray bars. Because the positions of the color items in the cue and search arrays were selected independently, color cues were uninformative with respect to the location of color bars in the search arrays. In the unimodal color task, participants had to respond to red bars and ignore blue bars, or vice versa (with target color varied across participants). No auditory stimuli were presented in this task. In the audiovisual color-pitch task, all search arrays were accompanied by a tone, and target trials were defined by a specific color-pitch combination. For example, participants were instructed to respond to bars when they were red (visual target-defining feature: V+) and were accompanied by a high-pitch tone (auditory target-defining feature: A+). In addition to these target trials (V+A+), there were also trials where one or both of these visual and auditory features did not match the current target definition (V- and A- trials, respectively). There were trials where a target-color bar was accompanied by a nontarget tone (e.g., red bar/low tone; V+A- trials), trials with a nontarget-color

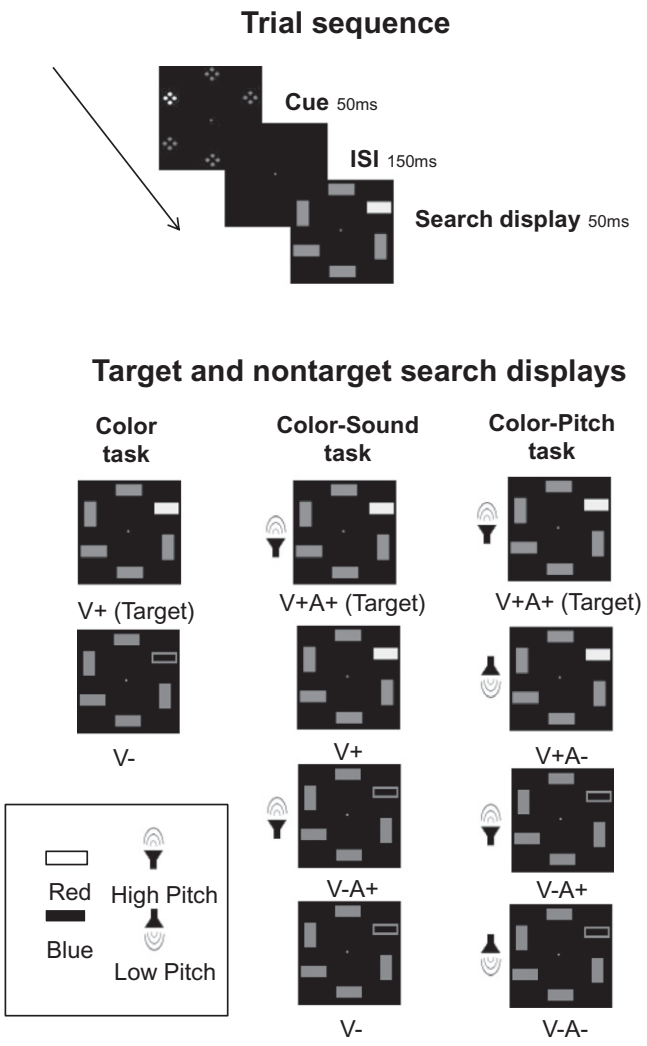


Figure 1. Top: Sequence of events on each trial. Bottom: Illustration of target and nontarget search arrays in the three task conditions of Experiment 1. In this example, a participant searched for red singleton bars (shown as white bars) and ignored blue bars (shown as black bars with gray outlines) in the color task. In the color-sound task, a red bar accompanied by a tone was a target, while red bars without tones and blue bars with or without tones had to be ignored. Red bars accompanied by high tones were targets in the color-pitch task, while other color/pitch combinations had to be ignored.

bar and a target-pitch tone (e.g., blue bar/high tone; V-A+ trials), and trials where both visual and auditory features were task-irrelevant (e.g., blue bar/low tone; V-A- trials). No response was required on any of these trials. In the audiovisual color-sound task, participants were instructed to respond on trials with target-color bars, but only when they are accompanied by a tone (V+A+ trials). They had to ignore target-color bars without tones (V+ trials), as well as all nontarget-color bars, regardless of whether they were presented with or without tone (V-A+ and V- trials). Thus, the two audiovisual tasks differed with respect to the auditory judgment required to distinguish target and nontarget trials. In the color-pitch task, a pitch discrimination (high vs. low tone) was needed. In the color-sound task, an auditory present/absent judgment was sufficient.

In all three tasks, cue arrays were physically identical and always included a task-set matching color singleton item (see Figure 1). In this respect, procedures differed from standard contingent attentional capture experiments, which typically employ both task-set matching and nonmatching cues (e.g., Eimer & Kiss, 2008; Folk et al., 1992). We presented only color singleton cues that matched the visual target-defining feature in order to find out whether the ability of these cues to attract attention differed between audiovisual and unimodal visual task contexts. In the unimodal visual task, color singleton cues should capture attention in a task-set contingent fashion, in spite of the fact that they conveyed no information with respect to the location of the color bar in the subsequent search array. Attentional capture by target-color matching singleton cues should be reflected by faster RTs to target bars at cued as compared to uncued locations (e.g., Folk et al., 1992), and by the presence of N2pc components in response to these singleton cues (e.g., Eimer & Kiss, 2008). The critical question was whether the same effects would also be observed during search for audiovisually defined targets. If audiovisual search was guided by strictly modality-specific attentional templates that operate independently in vision and audition, the fact that the color cues matched the currently task-relevant color should be sufficient to produce behavioral and electrophysiological attentional capture effects during audiovisual search that are identical to the effects triggered in the unimodal visual search task. In contrast, if the guidance of attention was integrated across modalities, a different pattern of results should be obtained. Because the color singleton cues were not accompanied by tones and therefore only partially matched the target features in the audiovisual tasks, their ability to capture attention should be reduced in these tasks relative to the unimodal visual task, and this should be reflected by smaller behavioral spatial cueing effects and reduced N2pc amplitudes. In the extreme case, these attentional capture effects should be entirely absent during audiovisual search.

Experiment 1

Method

Participants. Twelve right-handed paid volunteers with normal or corrected vision (mean age 25.8 years, age range 21–40 years, 5 females) took part. Informed consent was obtained from all participants prior to the start of the experiment.

Stimuli and apparatus. Visual stimuli were presented at a viewing distance of 100 cm on a 22″ LCD monitor (Samsung wide SyncMaster 2233; 100 Hz refresh rate) against a black background. On each trial, a cue display (50-ms duration) was followed after a 150-ms interstimulus interval by a search array (50-ms duration). Intertrial interval was 1,450 ms. Each cue and search array contained a circular array of six elements at a distance of 4.1° from a central fixation point (Figure 1, top panel). Cue arrays contained six elements composed of four closely aligned dots (0.17° × 0.17°). One set of dots was a color singleton that matched the target color (blue or red, varied across subjects; CIE [International Commission on Illumination] *x/y* chromaticity coordinates .161/.128 and .621/.128, respectively). This color singleton was presented equiprobably and randomly at one of the four lateral locations, but never at the top or bottom. The five remaining cue elements were uniformly gray (.308/.345). Search arrays contained six horizontal or vertical bars (1.1° × 0.3°) at the same positions as the preceding cue elements, with bar orientation chosen randomly for each position. One of these bars was colored (blue or red), the others were gray. Colored bars appeared with equal probability at one of the four

lateral locations. All gray, blue, and red stimuli in the cue and search displays were equiluminant (~11 cd/m^2). In two of the three search tasks, search arrays could be accompanied by auditory stimuli. These were pure sine-wave tones (50-ms duration; 65 dB SPL; high-pitch: 2000 Hz; low-pitch: 300 Hz) that were presented concurrently with search array onset from a loudspeaker located centrally behind the monitor.

Procedure. The experiment included three search task conditions. In all three tasks, color singleton cue arrays were identical, but targets were defined in a different way, and target as well as nontarget search arrays were also different (as illustrated in Figure 1, bottom panel). In the unimodal color task, participants had to respond to target-color bars (e.g., red bars), and to ignore nontarget-color bars (e.g., blue bars during search for red bars). Both trial types were presented with equal probability and in random order. In the audiovisual color-pitch task, all search arrays were accompanied by synchronous tones. Participants were instructed to respond to color singleton bars in the search array when they matched the current target color and were accompanied by a specific tone (e.g., red bars accompanied by high tones). On half of all trials, both target-defining visual and auditory features were present (V+A+ trials). On the other half of trials, one or both of these features were absent, and no response was required. On V+A- trials, there was a nonmatching auditory feature (e.g., red bar/low tone). On V-A+ trials, the visual feature did not match the target definition (e.g., blue bar/high tone). On V-A-, neither the visual nor the auditory feature matched the target (e.g., blue bar/low tone). These three nontarget trial types appeared with equal probability, and the assignment of target and nontarget tone frequencies was counterbalanced across participants. In the audiovisual color-sound task, participants were instructed to respond to target-color bars when they were accompanied by a tone, which was the case on half of all trial (V+A+ trials, see Figure 1). They had to ignore target-color bars that were presented without concurrent tones (V+ trials), and nontarget-color bars regardless of whether they appeared with or without a tone (V-A+ trials and V-A- trials). These three nontarget trial types were equiprobable.

Four successive blocks were run for each task, and each task run was preceded by two training blocks. Task order was counterbalanced across participants. Each block included 96 trials (48 target trials and 48 nontarget trials), resulting in a total number of 1,152 trials across all twelve experimental blocks (384 trials for each of the three task conditions). The assignments of target and nontarget features (color and pitch) remained constant across all three tasks for each participant, and were counterbalanced across participants. In all three tasks, participants responded to the orientation of this color singleton bar on target trials by pressing one of two vertically aligned response keys. Vertical and horizontal target bars were mapped to the top and bottom key, respectively. The assignment of the left or right hand to the top or bottom response key was counterbalanced across participants.

EEG recording and data analysis. Electroencephalogram (EEG) was DC-recorded from 23 scalp electrodes mounted in an elastic cap at standard positions of the extended 10–20 system at sites Fpz, Fz, F3, F4, F7, F8, FC5, FC6, Cz, C3, C4, T7, T8, CP5, CP6, Pz, P3, P4, P7, P8, PO7, PO8, and Oz (500 Hz sampling rate; 40 Hz low-pass Butterworth filter). All scalp electrodes were online referenced to the left earlobe and rereferenced offline to the average of both earlobes. Impedances were kept below 5 kΩ. Horizontal eye movements (HEOG) were measured from two electrodes placed at the outer canthi of the eyes. Only trials with correct responses to targets or correctly withheld responses to nontargets were analyzed. Trials with saccades (voltage exceeding ± 30 μV in the HEOG channel), eye blinks (exceeding ± 60 μV at Fpz), or muscle artifacts (exceeding ± 80 μV at any other electrode) were excluded from the analyses, as were trials with incorrect responses, missed targets, or false alarms.

EEG in response to cue stimuli was epoched and averaged for the 500-ms interval after cue onset, relative to a 100-ms precue baseline. Averages were computed for trials with color singleton cues in the left and right hemifield, separately for all three tasks. N2pc amplitudes were quantified on the basis of mean amplitudes obtained between 170 ms and 270 ms after cue onset at lateral posterior electrodes PO7 and PO8. Onset latencies of cue-elicited N2pc components were compared between the three search tasks, using the jackknife method described by Miller, Patterson, and Ulrich (1998). This method is based on N2pc difference waveforms obtained by subtracting ERPs at electrodes PO7/8 ipsilateral to the side of the color singleton cues from contralateral ERPs, which were computed separately for each individual participant and each search task. Next, a set of subsamples of grand-averaged N2pc difference waveforms was obtained by successively excluding one individual participant from the original sample, separately for each task. N2pc onset latency differences between task conditions were then estimated by comparing the sets of subsamples between tasks with paired *t* tests. N2pc onset latency was defined relative to an absolute amplitude criterion of −1 μV (see Eimer, Kiss, & Nicholas, 2011, for similar procedures), and *t* values were corrected according to the formula described by Miller et al. (1998). In all analyses, Greenhouse-Geisser corrections for violated sphericity assumptions were applied where appropriate.

Results

Behavioral performance. Trials with RTs below 200 ms or above 1,000 ms, or trials where RTs deviated by more than ±3 *SD*s from the mean were excluded from analyses (less than 1% of all trials). Figure 2 (left panel) shows RTs for correct responses to targets at cued and uncued locations, separately for the three search tasks. A repeated measures analysis of variance (ANOVA) was conducted on the RT data for the factors spatial cueing (target at cued vs. uncued location) and task (color, color-sound, color-pitch). There was no main effect of task, $F < 1$. A main effect of spatial cueing, $F(1,11) = 27.6$, $p < .001$, $\eta_p^2 = .715$, reflected faster RTs to targets at cued versus uncued locations, indicative of cue-induced attentional capture. Most importantly, a two-way interaction between spatial cueing and task was obtained, $F(2,22) = 9.4$, $p < .01$, $\eta_p^2 = .461$, demonstrating that attentional capture by target-color cues differed between the unimodal visual and the two audiovisual search tasks. A spatial cueing effect of 37 ms was observed in the unimodal color task, $F(1,11) = 46.5$, $p < .001$, $\eta_p^2 = .809$. In the audiovisual color-sound task, this attentional capture effect was completely eliminated (−1 ms; $F < 1$). A planned comparison via a *t* test confirmed that RT cueing effects in the color task were indeed reliably larger than in the color-sound task, $t(11) = 3.69$, $p < .01$. In the audiovisual color-pitch task, a significant spatial cueing effect of 25 ms was observed, $F(1,11) = 23.6$; $p < .001$, $\eta_p^2 = .68$. This attentional capture effect was reliably smaller than the effect observed in the unimodal color task, $t(11) = 2.4$, $p < .05$.

Response errors were more frequent to targets at uncued locations relative to cued targets (4.4% vs. 2.6%; $F(1,11) = 6.3$, $p < .05$, $\eta_p^2 = .46$). Error rates did not differ between tasks, $F < 1$, and the

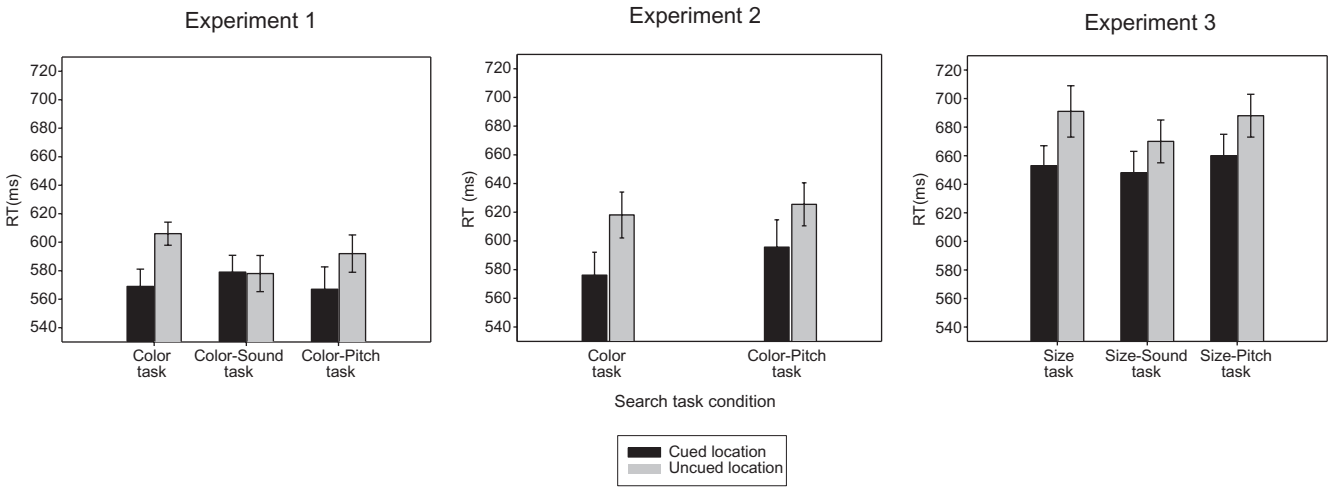


Figure 2. Mean correct RTs (in ms) to targets at cued and uncued locations, shown separately for each search task in Experiment 1 (left), Experiment 2 (middle), and Experiment 3 (right). Error bars represent standard errors of the mean.

interaction between spatial cueing and task was not significant, $F(2,22) = 1.7$, $p = .2$, $\eta_p^2 = .14$. Participants missed less than 1% of all targets on go trials. False alarms occurred on 1.2% of all no-go trials, and false alarm rates differed between the three search tasks, $F(2,22) = 6.2$, $p < .01$, $\eta_p^2 = .36$. False alarms were virtually absent in the unimodal color task (0.03%), and were relatively more frequent in color-sound and color-pitch tasks (1.4% and 2.1%, respectively). In these two tasks, false alarms were exclusively observed on nontarget trials with target-color bars (V+ trials or V+A- trials, respectively).

N2pc component. Figure 3 (top panels) shows ERPs triggered in response to cue arrays in the 350-ms interval after cue onset at electrodes PO7/8 contralateral and ipsilateral to the side of the color singleton cue, separately for the three search tasks. As expected, an N2pc component was triggered in response to target-color singleton cues in all three tasks. However, N2pc amplitudes and latencies differed between these tasks. This can be seen more clearly in the difference waveforms shown in Figure 3 (bottom panel), which were obtained by subtracting ipsilateral from contralateral ERPs: The N2pc appears to be larger and emerge slightly earlier in the unimodal color task relative to the two audiovisual tasks.

These differences were evaluated with a repeated measures ANOVA for the factors contralaterality (electrode ipsilateral vs. contralateral to the color singleton cue) and task (color, color-sound, color-pitch). A main effect of contralaterality on N2pc mean amplitudes, $F(1,11) = 21.7$, $p < .001$, $\eta_p^2 = .66$, was accompanied by an interaction between contralaterality and task, $F(2,22) = 3.8$, $p < .05$, $\eta_p^2 = .26$, demonstrating that N2pc amplitudes differed across the three tasks. Planned contrasts revealed that the N2pc elicited by cue arrays in the color-sound task was reliably smaller than the N2pc measured in the unimodal color task, $t(11) = 2.87$, $p < .01$ (see Figure 3, bottom panel). There was also a trend for a reduced N2pc in the color-pitch task relative to the color task, but this difference failed to reach significance, $t(11) = 1.28$, $p = .11$. The N2pc emerged significantly earlier in the unimodal color task (185 ms) than in the color-sound task (194 ms; $t_c(11) = 2.69$, $p < 0.05$, one-tailed). The N2pc onset latency difference between the color task and the color-pitch task (185 ms vs. 191 ms) failed to reach significance, $t_c(11) = 1.38$, $p = .098$.

The differences observed for cue-triggered N2pc components between the unimodal and the two audiovisual tasks could in principle be linked to the presence of additional auditory ERP components in the color-sound and color-pitch tasks, where some or all search arrays were accompanied by a tone. Even though these search arrays appeared 200 ms after the cue, and early auditory ERP components are typically not lateralized, the possibility that the presentation of tones produced an attenuation of N2pc components cannot be completely dismissed. To rule out this possibility, we compared cue-elicited N2pc amplitudes observed in the color-sound task on V+A+ trials (where cues were followed by a search array that contained a target-color bar and was accompanied by a sound) and on V+ trials (where the same search array was presented without concurrent sound). There was no evidence for any N2pc mean amplitude difference between V+A+ and V+ trials, $t(11) < 1$, demonstrating that the presence or absence of a subsequent tone did not affect the N2pc to the cue.

Discussion

To test whether attentional object selection during audiovisual search is guided by attentional templates that represent visual and auditory target features in an independent modality-specific fashion, or by integrated bimodal object representations, we compared attentional capture effects triggered by target-color matching visual singleton cues during search for unimodal visual targets and during search for audiovisually defined targets. Behavioral and electrophysiological results demonstrated that audiovisual search is not controlled in an exclusively modality-specific fashion, and that early stages of visual object selection are already differentially modulated during visual versus audiovisual search.

As expected, target-color matching singleton cues captured attention in the unimodal color task, as reflected by behavioral spatial cueing effects and N2pc components triggered by these cues. This is in line with previous behavioral and electrophysiological evidence for task-set contingent attentional capture (e.g., Eimer & Kiss, 2008; Folk et al., 1992). In spite of the fact that the

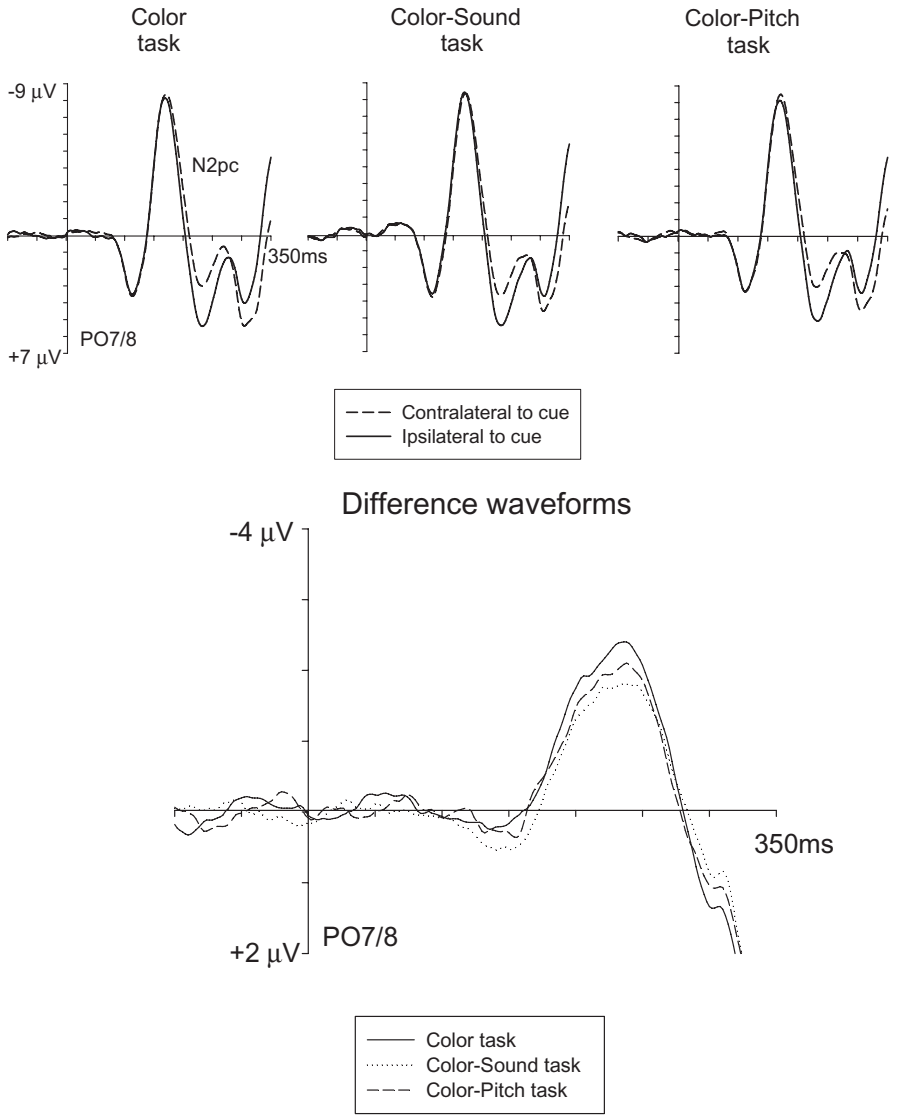


Figure 3. Top: Grand-average ERPs measured in Experiment 1 at posterior electrodes PO7/8 contralateral and ipsilateral to the location of a target-color singleton cue, separately for the color task, the color-sound task, and the color-pitch task. Bottom: Difference waveforms obtained by subtracting ipsilateral from contralateral ERPs, shown separately for the three search tasks.

color singleton cues were physically identical across all three tasks, behavioral attentional capture effects were substantially reduced in the two audiovisual tasks. In the color-sound task, RT spatial cueing effects were completely eliminated. In the color-pitch task, they were significantly reduced relative to the unimodal color task. Along similar lines, the N2pc triggered by color singleton cues was significantly reduced in amplitude and delayed in the color-sound task as compared to the color task. Trends toward a reduction and delay of the N2pc were also observed for the color-pitch relative to the unimodal color task, although they did not reach statistical significance. If attentional selectivity in the two audiovisual tasks had been guided by strictly separate modality-specific representations of task-relevant visual and auditory features, behavioral and electrophysiological correlates of attentional capture by color-singleton cues in these tasks should have been identical to the effects observed in the unimodal color task, as these singleton cues always matched the target-defining color. The observation that behavioral spatial cueing effects were reduced in the color-pitch task and entirely absent in the color-sound task, and the fact that the N2pc to color singleton cues was attenuated and delayed in the color-sound task, strongly suggest that the ability of these cues to capture attention was reduced during audiovisual search. These observations point to an important role for integrated bimodal object templates in the control of search for audiovisually defined targets.

In the color-sound task, behavioral and electrophysiological markers of attentional capture showed a surprising dissociation. Behavioral spatial cueing effects were completely absent in this task, suggesting that target-color singleton cues failed to capture attention. However, although the N2pc component was attenuated relative to unimodal color search, it remained reliably present, indicating that target-color cues retained some of their ability to attract attention. This difference between electrophysiological and behavioral measures suggests that they reflect different aspects of task-set contingent attentional capture. We will return to this issue in the general discussion.

While behavioral and electrophysiological markers of attentional capture were reliably reduced in the audiovisual color-sound task relative to the unimodal color task, the attenuation of cue-induced capture effects was less pronounced in the color-pitch task, where the N2pc reduction only approached statistical significance. Why was attentional capture by visual singleton cues more strongly reduced in the color-sound task? Task difficulty is unlikely to be a major factor, as target detection performance did not differ between the color-sound and color-pitch tasks, even though these two tasks required different auditory discriminations (tone detection vs. pitch discrimination). A more plausible candidate is the perceptual similarity between cue arrays and some of the nontarget search arrays in the color-sound task. This task included trials where search arrays with a target-color singleton bar were presented without a synchronous sound (V+ trials, see Figure 1), and no response was required on these trials. Because these search arrays were designated as nontargets, participants may have adopted a top-down inhibitory attentional set towards them. As singleton cue arrays were perceptually similar to the search arrays on V+ trials (i.e., both included a target color-matching singleton without a concurrent tone), such an inhibitory set may also have been applied to these cue arrays in the color-sound task, resulting in a reduction of attentional capture in this task. In contrast, all search arrays were accompanied by tones in the color-pitch task. The fact that cue arrays did not perceptually match any of the nontarget arrays in this task may have resulted in less inhibition of attentional capture. While this explanation can account for the differences between the two audiovisual tasks in Experiment 1, the fact remains that electrophysiological attentional capture effects did not differ reliably between the unimodal color task and the audiovisual color-pitch task. This may cast doubt on the hypothesis that integrated bimodal attentional templates play an important role in the guidance of search for audiovisual targets. In Experiment 2, participants were given a stronger incentive to process the auditory target-defining attribute in the color-pitch task.

Experiment 2

To facilitate participants' focus on the auditory aspects of the color-pitch task, the task relevance of the target pitch was increased in Experiment 2 by changing the proportion of nontarget trial types (see Bacon & Egeth, 1997, for evidence that distractor probabilities affect top-down search strategies during search for conjunctively defined visual targets). Relative to Experiment 1, the number of nontarget trials where a target-pitch tone was presented simultaneously with a nontarget-color bar (V-A+ trials) was reduced from 16 to 4, while the number of trials with target-color bars and nontarget-pitch tones (V+A- trials) was increased from 16 to 28. As a result, the presence of the target-pitch sound was now much more strongly associated with the target status of a given trial. This manipulation should result in the auditory target-defining feature being weighted more strongly in the color-pitch task, and thus in a reliable reduction of both behavioral and electrophysiological markers of attentional capture by unimodal visual cues relative to the color task.

Method

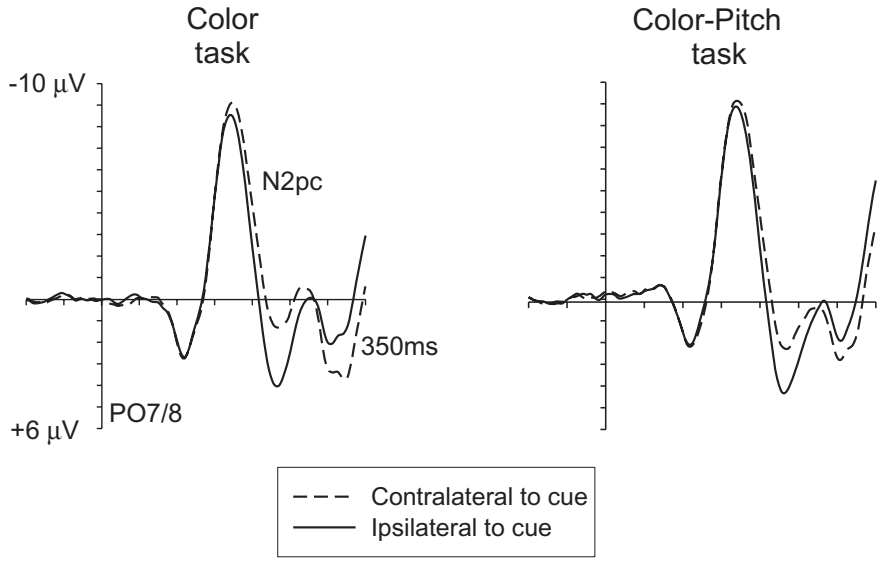
Participants. Thirteen paid volunteers took part in the study. One participant was excluded due to excessive EEG activity in the alpha band. The remaining 12 participants (mean age 28.5 years, age range 22–38 years; 1 left-handed; 5 males) had normal or corrected-to-normal vision. All gave informed consent to participate in the study.

Stimuli, apparatus, and procedure. Experimental setup and procedures were identical to Experiment 1 with the following exceptions. Only the unimodal color task and the audiovisual color-pitch task were included. Four blocks of 96 trials were run for each task, resulting in a total number of 768 trials across all eight experimental blocks (384 trials for each task). In the color-pitch task, the ratio of V+A- and V-A+ trials was altered in order to make target-matching tones more strongly indicative of the presence of a response-relevant target, and to encourage participants to focus on both auditory and visual target features during audiovisual search. In Experiment 1, each block included 16 V-A+ trials and 16 V+A- nontarget trials. In Experiment 2, there were only 4 V-A+ trials and 28 V+A- trials in each block. The number of V+A+ and V-A- trials per block (48 vs. 16) remained unchanged. This change implied that the presence of the target-defining pitch (A+) on any given trial was now more strongly associated with the target status of this trial: On 92% of all trials where a target-matching pitch was present (48 out of 52 trials per block), participants had to discriminate and respond to the target-color bar. In contrast, the strength of the association between the presence of the target color (V+) and the target status of a given trial was now reduced: There were 76 trials where a target-color bar was present, but only 48 of these (63%) required a response. Participants were informed that trials where a target pitch was accompanied by a search array with a nontarget-color bar were quite rare (but their exact number was not revealed), and that search would be facilitated by preparing specifically for the specific combination of target pitch and target color. To prevent them from adopting a unimodal auditory task set, participants were explicitly instructed not to respond on those rare trials where a V-A+ stimulus was presented.

EEG recording and data analysis. These were identical to Experiment 1, except that task was now a two-level factor (color vs. color-pitch task).

Results

Behavioral performance. Trials with anticipatory and exceedingly slow responses (defined as in Experiment 1) were excluded, resulting in a loss of less than 1% of all trials. Figure 2 (middle panel) shows RTs for correct responses to targets presented at cued and uncued locations, separately for two search tasks. A two-way ANOVA with spatial cueing and task as within-subject factors revealed a trend for responses to be faster in the unimodal color task (597 ms) than in the audiovisual color-pitch task (610 ms), $F(1,11) = 4.25$, $p = .064$, $\eta_p^2 = .28$. As in Experiment 1, there was a main effect of spatial cueing, $F(1,11) = 64.51$, $p < .001$, $\eta_p^2 = .85$, with faster RTs to cued targets, indicative of cue-induced attentional capture. A significant spatial cueing effect of 42 ms was observed in the color task, $F(1,11) = 95.1$, $p < .001$, $\eta_p^2 = .9$, and this effect was smaller in size (30 ms) but still significant in the color-pitch task, $F(1,11) = 23.74$, $p < .001$, $\eta_p^2 = .68$. There was a reliable interaction between task and spatial cueing, $F(1,11) = 4.98$, $p < .05$, $\eta_p^2 = .31$, demonstrating that the ability of target-color cues to capture attention was reduced in the color-pitch task as compared to the unimodal color task. Incorrect responses were more frequent for uncued as compared to cued targets (1.9% vs. 0.7%,



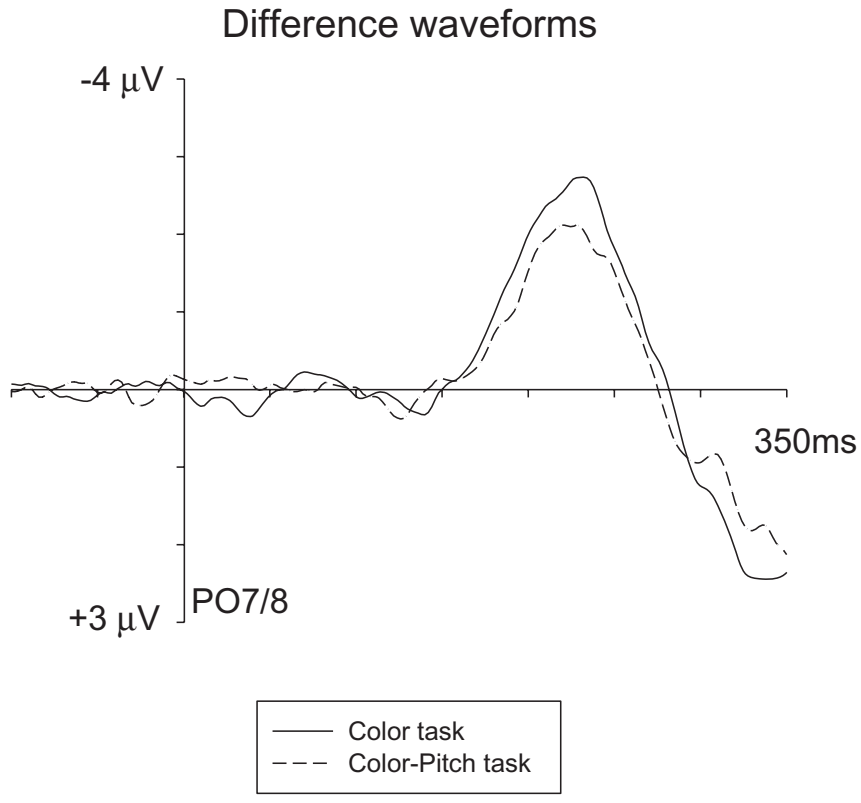


Figure 4. Top: Grand-average ERPs measured in Experiment 2 at posterior electrodes PO7/8 contralateral and ipsilateral to the location of a target-color singleton cue, separately for the color task and the color-pitch task. Bottom: Difference waveforms obtained by subtracting ipsilateral from contralateral ERPs, shown separately for the two search tasks.

$F(1,11) = 4.89$, $p < .05$, $\eta_p^2 = .31$). There was no main effect of task on error rates, $F < 1$, and no interaction between task and spatial cueing, $F(1,11) = 2.5$, $p = .14$, $\eta_p^2 = .19$. Participants failed to respond on less than 1% of all target trials. False alarms occurred on 0.3% of all nontarget trials.

N2pc component. Figure 4 (top panels) shows ERPs triggered in response to cue arrays in the color and color-pitch tasks at electrodes PO7/8 contralateral and ipsilateral to the side of the target-color singleton cue. An N2pc component was triggered by these singleton cues in both tasks. However, as can be seen in the

contralateral-ipsilateral difference waveforms shown in Figure 4 (bottom panel), N2pc amplitude was reduced, and N2pc onset was delayed in the color-pitch task relative to the unimodal color task. This was confirmed by statistical analyses. For N2pc mean amplitudes measured in the 170–270 ms postcue time window, there was a main effect of contralaterality, $F(1,11) = 43.37$, $p < .001$, $\eta_p^2 = .80$, and follow-up analyses revealed that a reliable N2pc was triggered by target-color cues in the unimodal color task, $F(1,11) = 44.41$, $p < .001$, $\eta_p^2 = .80$, as well as in the audiovisual color-pitch task, $F(1,11) = 35.87$, $p < .001$, $\eta_p^2 = .77$. Critically, there was now also a reliable interaction between contralaterality and task, $F(1,11) = 9.68$, $p < .01$, $\eta_p^2 = .47$, demonstrating that the N2pc was reduced in amplitude in the color-pitch task relative to the color task. The analysis of N2pc onset latencies revealed that the N2pc to target-color cues emerged reliably earlier in the unimodal color task than in the audiovisual color-pitch task (181 ms vs. 191 ms, $t_c(11) = 1.84$, $p < .05$, one-tailed).

Discussion

The task relevance of target-defining auditory features in the color-pitch task was increased relative to Experiment 1 by changing the relative probability of nontarget trial types, so that the presence of the target-defining pitch was now more strongly associated with the target status of a given trial. Reliable reductions of behavioral as well as electrophysiological markers of attentional capture by target-color cues were observed during audiovisual as compared to unimodal visual search. RT spatial cueing effects were significantly smaller in the color-pitch task than in the unimodal color task. In contrast to Experiment 1, the N2pc to target-color singleton cues was now reliably attenuated and delayed in the color-pitch task as compared to the unimodal color task. These findings confirm that bimodal attentional templates play an important role in the guidance of search for audiovisual targets. However, although behavioral and ERP markers of attentional capture were reliably attenuated in the color-pitch task, they were still significant, which suggests that unimodal target-color cues retained some of their ability to attract attention in this audiovisual task context. This issue will be further considered in the general discussion.

Experiment 3

Experiments 1 and 2 demonstrated that task-set contingent attentional capture by color singleton cues is reduced during search for audiovisual as compared to purely visually defined targets, suggesting that bimodal attentional templates are active during audiovisual search, and that these templates modulate the selective processing of visual stimuli in modality-specific extrastriate cortical areas. Experiment 3 investigated whether the control of visual selection by bimodal attentional templates is specific to situations where color is the target-defining visual attribute, or can also be found for other task-relevant visual dimensions. We tested whether attentional capture by visual singleton cues is attenuated in a bimodal task context where the task-relevant visual dimension is size. Procedures were identical to Experiment 1, except that color singletons were replaced by size singletons in the cue and search arrays. In the unimodal size task, participants had to discriminate the orientation of small singleton bars among medium-size distractors, and ignore search arrays with large singleton bars. The two audiovisual size-sound and size-pitch tasks were identical to the color-sound and color-pitch tasks of Experiment 1, except that red and blue bars were now replaced by small and large bars. Small bars were task relevant, and large bars had to be ignored. In all three tasks, search arrays were preceded by spatially uninformative target-matching (small) size singleton cues. If search for audiovisually defined targets was generally guided by integrated bimodal templates, attentional capture by target-matching size singleton cues should again be attenuated in the audiovisual tasks relative to the unimodal visual size task. In contrast, if the results observed in Experiments 1 and 2 were specific to tasks with color-defined targets, behavioral and electrophysiological attentional capture effects should be equivalent across all three tasks, demonstrating that search for size/sound and size/pitch targets is controlled in an independent modality-specific fashion.

Method

Participants. Sixteen paid volunteers were tested. Three were excluded due to excessive eye movements, and another due to an inability to discriminate between size-defined visual targets and nontargets. The 12 remaining participants (mean age 27.9 years, age range 22–42 years, 6 females) were all right-handed and had normal or corrected vision. All gave informed consent prior to participation.

Stimuli, apparatus, and procedure. Experimental setup and procedures were the same as in Experiment 1, except that size now replaced color as the visual feature singleton dimension. Cue arrays contained one smaller set of dots ($0.11° \times 0.11°$) among five larger sets ($0.17° \times 0.17°$). Search arrays always contained a one-size singleton bar (small: $0.7° \times 0.17°$, or large: $1.9° \times 0.57°$) among five medium-size bars ($1.1° \times 0.3°$). All visual stimuli were gray (CIE *x/y* coordinates: .308/.345; luminance: 11 cd/m^2). For all participants, small bars served as visual target-defining stimuli (V+), while large bars served as visual nontargets (V-). As size instead of color was now used as the visual target-defining dimension, the three tasks performed by the participants were now termed size task, size-sound task, and size-pitch task. The structure and trial probabilities for each of these three tasks were identical to the color, color-sound, and color-pitch tasks of Experiment 1.

EEG recording and data analysis. These were identical to Experiment 1, except that different time windows and onset criterion values were used for the N2pc analyses. This was due to the fact that N2pc components to small-size singleton cues were considerably smaller and emerged later than the N2pc triggered by target-color singleton cues in Experiments 1 and 2. N2pc mean amplitudes were now measured during the 200–310 ms interval after cue onset, and an absolute amplitude criterion of –0.4 μV was used for the jackknife-based analyses of N2pc onset latencies.

Results

Behavioral performance. Exclusion of trials with anticipatory and very slow responses (defined as in Experiments 1 and 2) led to a loss of 1.2% of all data. Figure 2 (right panel) shows RTs for correct responses to targets at cued and uncued locations, separately for the three search tasks. There was no main effect of task on RTs, $F < 1.5$. A main effect of spatial cueing, $F(1,11) = 30$, $p < .001$, $\eta_p^2 = .73$, reflected faster RTs to targets at cued versus uncued locations, indicative of attentional capture by size singleton cues. Critically, there was an interaction between spatial cueing and task, $F(2,22) = 5.9$, $p < .01$, $\eta_p^2 = .35$, demonstrating differential

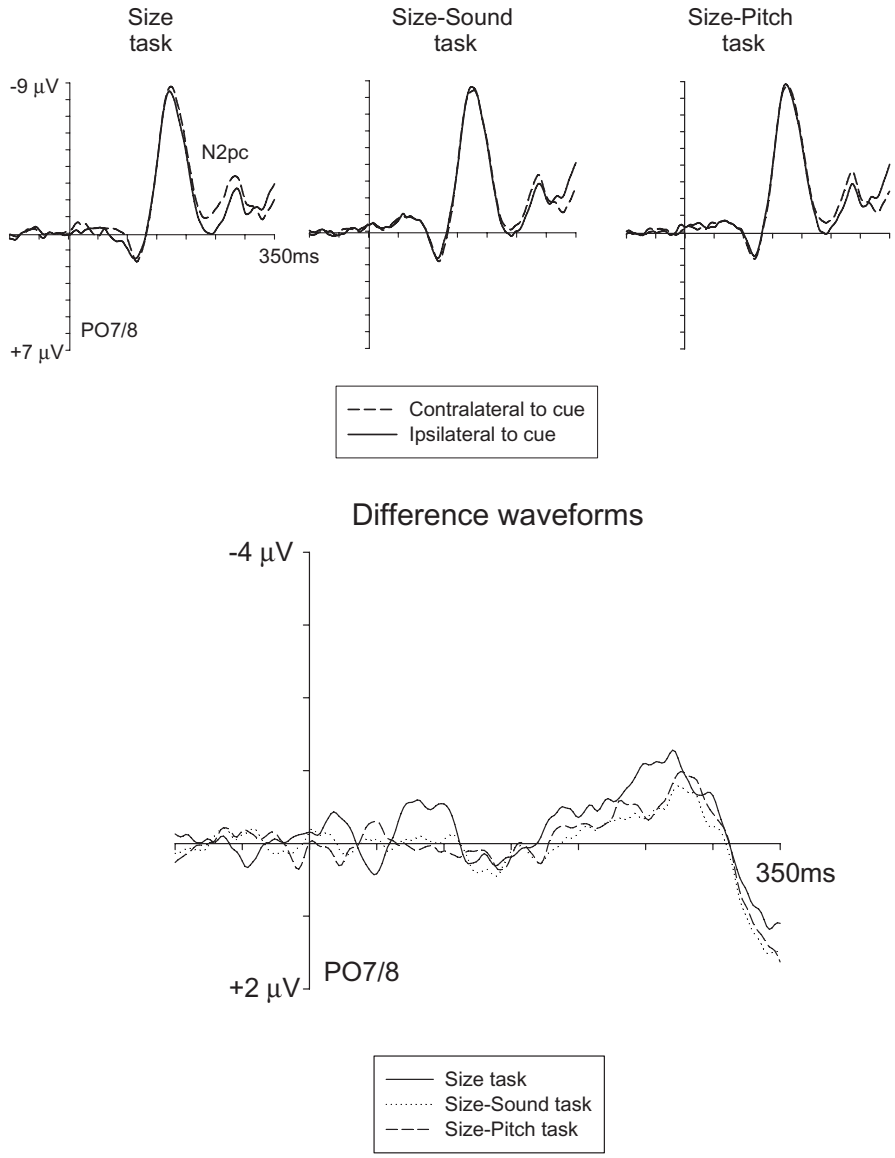


Figure 5. Top: Grand-average ERPs measured in Experiment 3 at posterior electrodes PO7/8 contralateral and ipsilateral to the location of a target-size (small) singleton cue, separately for the size task, the size-sound task, and the size-pitch task. Bottom: Difference waveforms obtained by subtracting ipsilateral from contralateral ERPs, shown separately for the three search tasks.

attentional capture by size singleton cues across the three tasks. A spatial cueing effect of 38 ms in the unimodal size task was reduced to 22 ms and 28 ms in the audiovisual size-sound and size-pitch tasks, respectively. These spatial cueing effects were significant in all three tasks, all $F(1,11) > 15.0$, all $p < .01$. Planned comparisons via one-tailed t tests revealed that cue-induced attentional capture effects on RTs were reliably larger in the unimodal size task than in the size-sound task, $t(11) = 3.18$, $p < .01$, and in the size-pitch task, $t(11) = 2.16$, $p < .05$.

Response errors were more frequent to uncued relative to cued targets (9.6% vs. 5.4%; $F(1,11) = 11.8$, $p < .01$, $\eta_p^2 = .52$). There was no main effect of task on error rates, and no interaction between spatial cueing and task, both $F < 1$. Participants missed 3% of all targets on go trials. False alarms occurred on approximately 2% of all nontarget trials, and false alarm rates did not differ significantly between the three search tasks. In both audiovisual tasks, false alarms were more frequent on nontarget trials with small (V+) bars than on nontarget trials with large (V-) bars.

N2pc component. Figure 5 (top panels) shows ERPs triggered in response to cue arrays with small size singleton cues at PO7/8 contralateral and ipsilateral to the side of these cues, separately for the three search tasks. Although N2pc components triggered by the size singleton cues were considerably smaller than the N2pc components elicited by color singleton cues in Experiments 1 and 2, they were present in all three tasks. As can be seen in the contralateral-ipsilateral difference waveforms shown in Figure 5 (bottom panel), the N2pc elicited by size singleton cues in the unimodal size task was larger and emerged earlier than the N2pc measured in the two audiovisual tasks. This was confirmed by statistical analyses. For N2pc mean amplitudes measured in the 200–310 ms postcue interval, a main effect of contralaterality,

$F(1,11) = 15$, $p < .01$, $\eta_p^2 = .58$, was accompanied by an interaction between contralaterality and task, $F(2,22) = 7.4$, $p < .01$, $\eta_p^2 = .41$, demonstrating that the size of the N2pc varied across the three search tasks. Planned contrasts revealed that N2pc amplitudes were significantly larger in the unimodal size task relative to the audiovisual size-sound and size-pitch tasks, $t(11) = 3.17$, $p < .01$, and $t(11) = 2.5$, $p < .05$, respectively. Analyses of N2pc onset latencies confirmed that the N2pc emerged earlier in the unimodal size task (203 ms) than in the size-sound task (255 ms; $t_c(11) = 2.64$, $p < .05$), and in the size-pitch task (226 ms, $t_c(11) = 4.32$, $p < .01$).

Discussion

The ability of visual singleton cues to attract attention is reduced during audiovisual search as compared to unimodal visual search, and this reduction is not confined to tasks where color is the task-relevant visual dimension, but can also be found during search for size-defined targets. As observed in Experiments 1 and 2 for target-color singleton cues, target-matching small singleton cues elicited significantly smaller RT spatial cueing effects indicative of attentional capture in the two audiovisual tasks than in the unimodal visual task. Also confirming the observations from the first two experiments, the N2pc to size singleton cues was attenuated and delayed in the two audiovisual tasks relative to the unimodal task. If target-defining visual and auditory features had been represented independently in a strictly modality-specific fashion, very similar behavioral and electrophysiological markers of attentional capture should have been observed across all three tasks. The presence of systematic differences in cue-induced capture between the visual and the two audiovisual tasks thus provides further evidence that bimodal attentional templates play an important role in the guidance of search for targets that are defined by a combination of features from different sensory modalities.

It should be noted that the attentional capture effects observed in Experiment 3 were based on a task set for relative rather than absolute size. Because all cue array elements were much smaller than the small, medium, or large bars in the search arrays, the ability of small singleton cues to attract attention when participants searched for small target bars was not due to the fact that they matched the absolute size of the target bars. Instead, they captured attention because their relative size in their perceptual context matched the relative size of targets among distractor bars in search arrays (see also Kiss & Eimer, 2011, for analogous results in a study of task-set contingent attentional capture by size singletons, and Becker, Folk, & Remington, 2010, for a more general investigation of the role of relational information in contingent capture).

There were also some differences between the results obtained in Experiment 3 for size-defined singleton stimuli, and in Experiments 1 and 2 where color singletons were employed. The N2pc triggered by size singleton cues was much smaller than the N2pc elicited by color singleton cues in the first two experiments. This is in line with previous ERP studies of attentional capture by visual feature singletons, which generally found larger N2pc components for color singletons as compared to singletons that are defined in another dimension such as shape (Seiss, Kiss, & Eimer, 2009) or size (Kiss & Eimer, 2011). It is possible that this N2pc amplitude difference between color singletons and other types of visual feature singletons reflects the stronger bottom-up salience of feature contrasts in the color domain. In spite of the N2pc amplitude differences between color and size singleton cues, the amplitude reductions and onset latency delays observed in audiovisual as compared to unimodal visual task contexts were very similar for both types of cues. This suggests that the impact of bimodal templates on attentional object selection is not affected by differences in the salience of the visual target-defining dimension.

Another difference between Experiments 1 and 3 concerns the pattern of behavioral spatial cueing effects. In Experiment 1, these effects were completely eliminated in the color-sound task as compared to the unimodal color task. In Experiment 3, they were significantly reduced in the analogous size-sound task, but remained reliably present. The fact that search for size-defined targets was more difficult than search for color-defined targets in Experiments 1 and 2 (as reflected by longer RTs and higher error rates in Experiment 3) may have been responsible for this difference. Reduced forward masking by size singleton cues could also have contributed to the residual spatial cueing effects in the size-sound task: Discriminating the orientation of small target singleton bars is likely to have been easier when it appears at a cued location (i.e., a location previously occupied by the smallest element of the cue array) than at uncued locations (i.e., locations previously occupied by a larger cue array element; see also Kiss & Eimer, 2011). However, as these factors were constant across all three tasks in Experiment 3, the reduction of spatial cueing effects for audiovisual as compared to unimodal task contexts can still be attributed to the impact of bimodal attentional templates.

Summary and Concluding Discussion

The aim of the present study was to investigate whether search for audiovisually defined targets is controlled in a modality-specific fashion by attentional templates that independently represent task-relevant visual and auditory features, or by integrated bimodal templates. Behavioral and electrophysiological markers of attentional capture by visual feature singleton cues that matched the currently task-relevant visual feature were measured in unimodal visual search tasks where targets were defined by this feature alone, and in audiovisual search tasks where targets were defined by a combination of visual and auditory features. If the guidance of attention by target-defining visual and auditory attributes operated strictly independently, attentional capture by task-set matching visual feature singleton cues should be identical during search for visual and for audiovisual targets. If integrated bimodal attentional templates were involved in the control of search for audiovisual targets, capture by visual cues should be attenuated or perhaps even completely eliminated in audiovisual task contexts.

Results demonstrated that top-down attentional control during audiovisual search is not implemented in a strictly modality-specific fashion. Attentional templates for visual and auditory target features do not operate entirely independently, and selective attentional processing in visual cortex is not guided exclusively by visual target-defining features. Regardless of whether color (Experiments 1 and 2) or size (Experiment 3) served as the visual target dimension, behavioral spatial cueing effects indicative of attentional capture were smaller during search for audiovisual as compared to unimodal visual targets, and the N2pc component to task-set matching visual singleton cues was attenuated and delayed during audiovisual search. These findings demonstrate that the ability of these cues to attract attention is reduced when participants search for targets that are defined by a combination of visual and auditory features, thus providing novel evidence for an important role of bimodal object templates in control of audiovisual search.

However, the role of bimodal attentional templates during audiovisual search is not absolute. If search for audiovisual targets

was entirely controlled by templates where visual and auditory target features are fully integrated into a single bimodal object, neither feature should be able to attract attention when presented in isolation. In this case, one would expect no attentional capture by visual target-matching cues during audiovisual search at all, as attention would only be allocated to fully template-matching objects. In fact, although behavioral and electrophysiological attentional capture effects were attenuated during audiovisual search in all three experiments, they were clearly not completely eliminated. These observations are not in line with attentional guidance by fully integrated bimodal objects, and suggest that attentional control of audiovisual search retains some modality-specific aspects. The fact that cue arrays always contained a singleton that matched the current visual target feature may have contributed to the robust presence of cue-elicited N2pc components in both visual and audiovisual tasks, as participants may have used the cue as a reminder of the task-relevant visual attribute. However, the observation that behavioral and electrophysiological markers of attentional capture in response to the same visual singleton cues were consistently reduced in audiovisual as compared to unimodal visual task contexts demonstrates that bimodal attentional templates are also involved in the guidance of search for audiovisual targets.

The hypothesis that the control of audiovisual search has both modality-specific and crossmodal aspects may also account for the surprising dissociation between behavioral and electrophysiological markers of attentional capture observed in the color-sound task of Experiment 1. In this task, behavioral spatial cueing effects indicative of attentional capture were completely absent, but task-set matching visual singleton cues still triggered reliable N2pc components, indicating that these cues retained some of their ability to capture attention. This pattern of results suggests that N2pc and RT effects reflect different aspects of task-set contingent attentional capture. An analogous dissociation was observed in our lab in a recent study of attentional capture during search for conjunctively defined visual targets (Kiss, Grubert, & Eimer, 2013). In these experiments, participants searched for target singleton bars that were defined by a specific combination of color (e.g., red) and size (e.g., small). Target arrays were preceded by cue arrays that contained a spatially uninformative color/size singleton that could have both, one, or neither of the two visual target features. Singleton cues that only matched one of these two target features failed to trigger RT spatial cueing effects indicative of attentional capture, suggesting that conjunction search was controlled by integrated object representations, and not by independent attentional templates for each target feature. However, partially target-matching singleton cues did trigger reliable N2pc components, in line with feature-specific rather than object-based attentional guidance. To reconcile these apparently contradictory findings, we suggested a two-stage account: During search for conjunctively defined visual targets, each target feature initially triggers rapid attentional capture, reflected by an N2pc, irrespective of whether other target-defining features are also present. However, attention is then rapidly disengaged from only partially target-matching stimuli, which results in the absence of behavioral spatial cueing effects for partially matching singleton cues. In other words, attentional guidance by independent features and guidance by integrated object representations reflect two separable and successive stages in the attentional selection of conjunctively defined targets. A similar account may also apply to search for audiovisually defined targets. The earliest stage of attentional target selection (reflected by the N2pc) may operate largely independently for visual and auditory features, while a second stage (reflected by behavioral spatial cueing effects) is guided by integrated audiovisual templates. This explanation can account for the remarkable dissociation of electrophysiological and behavioral correlates of attentional capture observed in the color-sound task of Experiment 1. However, the fact that cue-triggered N2pc components were consistently attenuated and delayed during audiovisual as compared to unimodal visual search demonstrates that even the early stage of attentional object selection that is reflected by the N2pc is not entirely under modality-specific top-down control.

The apparent duality of crossmodal and modality-specific aspects in the guidance of audiovisual search may reflect the dual nature of the underlying control mechanisms: Bimodal attentional templates are likely to be represented in multimodal brain areas, but affect spatially selective attentional processes that operate in modality-specific sensory cortical regions. With respect to the locus of bimodal templates, it is generally assumed that attentional templates are working memory representations (Carlisle et al., 2011; Desimone & Duncan, 1995; Olivers & Eimer, 2011). Although models of working memory often postulate modality-specific subsystems for the storage and maintenance of visual and auditory information (e.g., Baddeley, 1998), there is converging evidence that relevant visual and auditory features are both represented in multimodal brain areas. For example, dorsolateral prefrontal cortex (DLPFC) contains neurons that maintain integrated audiovisual representations of color and pitch (Fuster, Bodner, & Kroger, 2000). Superior temporal sulcus (STS) contributes to the maintenance and integration of object features from different modalities during audiovisual object categorization (Werner & Noppeney, 2010), and is also a neural substrate for crossmodal effects in spatial attention (McDonald, Teder-Sälejärvi, Di Russo, & Hillyard, 2003). The lateral intraparietal area (LIP) contains multisensory spatial maps (see Stein & Stanford, 2008, for a review), and is involved in the control of space-based and feature-based visual attention (Assad & Maunsell, 1995; Shulman, D'Avossa, Tansy, & Corbetta, 2002). Each of these areas might be involved in the representation of bimodal attentional templates that guide audiovisual search.

Although bimodal attentional templates may be represented in multimodal brain regions such as DLPFC, STS, or LIP, they will ultimately affect the operation of selective attention in modality-specific cortical areas. The N2pc in particular is a modality-specific visual component that is triggered by task-relevant visual stimuli and originates primarily from extrastriate ventral visual cortex (Hopf et al., 2000). The fact that the N2pc to task-set matching visual singleton cues component was attenuated and delayed during audiovisual as compared to purely visual search provides the first evidence that spatially selective attentional processes in modality-specific visual areas can be modulated by bimodal attentional templates for target objects that are defined by a combination of features from different modalities. How could this top-down control by bimodal templates be implemented? In the Guided Search model of visual object selection (Wolfe, 1994, 2007), the input from visual feature channels to the central salience map is weighted in accordance with the task relevance of these features. Because current target features are weighted strongly, they create a strong spatial bias in the activity profile on the saliency map, which results in the preferential attentional selection of target objects. It is plausible to assume that during search for audiovisual targets, target-defining visual attributes also receive a positive weighting, but that these top-down weights and the resulting spatial bias in favor of task-set

matching visual features are reduced when targets are defined across sensory modalities relative to unimodal single-feature search tasks. For example, during search for red singleton bars that are accompanied by high-pitch tones, feature channels that code red objects are less strongly weighted (and thus have less impact on the activity profile of the salience map) than during unimodal search for red bars. As a result, attentional capture by red singletons is reduced, as was indeed observed in the current study. Any top-down bias in favor of red stimuli might be further reduced in a task context where unimodal red singletons serve as to-be-ignored nontargets (such as in the color-sound task of Experiment 1). In summary, bimodal attentional templates are likely to be represented in multimodal brain regions, but affect modality-specific mechanisms of attentional object selection. This might account for the current pattern of findings, which indicate that top-down attentional guidance of audiovisual search has bimodal as well as modality-specific aspects.

Overall, the current results provide novel evidence that search for audiovisual target objects is not exclusively controlled by independently operating modality-specific representations of target-defining features. Bimodal attentional templates are involved in the guidance of audiovisual search, and these bimodal templates already affect early stages of attentional selectivity in extrastriate visual cortex.

References

Assad, J. A., & Maunsell, J. H. R. (1995). Neural correlates of inferred motion in primate posterior parietal cortex. *Nature*, *373*, 518–521. doi: 10.1038/373518a0

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception*, *55*, 485–496. doi: 10.3758/BF03205306

Bacon, W. F., & Egeth, H. E. (1997). Goal-directed guidance of attention: Evidence from conjunctive visual search. *Journal of Experimental Psychology: Human Perception & Performance*, *23*, 948–961. doi: 10.1037/0096-1523.23.4.948

Baddeley, A. (1998). Recent developments in working memory. *Current Opinion in Neurobiology*, *8*, 234–238. doi: 10.1016/S0959-4388(98)80145-1

Becker, S. I., Folk, C. L., & Remington, R. W. (2010). The role of relational information in contingent capture. *Journal of Experimental Psychology: Human Perception & Performance*, *36*, 1460–1476. doi: 10.1037/a0020370

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional templates in visual working memory. *The Journal of Neuroscience*, *31*, 9315–9322. doi: 10.1523/JNEUROSCI.1097-11.2011

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

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*, 501–517. doi: 10.1037/0096-3445.113.4.501

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458. doi: 10.1037/0033-295X.96.3.433

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology*, *99*, 225–234. doi: 10.1016/0013-4694(96)95711-9

Eimer, M., & Kiss, M. (2008). Involuntary attentional capture is determined by task set: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *20*, 1423–1433. doi: 10.1162/jocn.2008.20099

Eimer, M., Kiss, M., & Nicholas, S. (2011). What top-down task sets do for us: An ERP study on the benefits of advance preparation in visual search. *Journal of Experimental Psychology: Human Perception & Performance*, *37*, 1758–1766. doi: 10.1037/a0024326

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 & Performance*, *35*, 1316–1328. doi: 10.1037/a0015872

Eimer, M., van Velzen, J., & Driver, J. (2002). Cross-modal interactions between audition, touch, and vision in endogenous spatial attention: ERP evidence on preparatory states and sensory modulations. *Journal of Cognitive Neuroscience*, *14*, 254–271. doi: 10.1162/089892902317236885

Folk, C. L., & Remington, R. W. (1998). Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception & Performance*, *24*, 847–858. doi: 10.1037/0096-1523.24.3.847

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 & Performance*, *18*, 1030–1044. doi: 10.1037/0096-1523.18.4.1030

Fuster, J. M., Bodner, M., & Kroger, J. K. (2000). Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature*, *405*, 347–351. doi: 10.1038/35012613

Hickey, C., McDonald, J. J., & Theeuwes, J. (2006). Electrophysiological evidence of the capture of visual attention. *Journal of Cognitive Neuroscience*, *18*, 604–613. doi: 10.1162/jocn.2006.18.4.604

Hopf, J.-M., Luck, S. J., Girelli, M., Mangun, G. R., Scheich, H., & Heinze, H.-J. (2000). Neural sources of focused attention in visual search. *Cerebral Cortex*, *10*, 1233–1241. doi: 10.1093/cercor/10.12.1233

Iordanescu, L., Grabowecky, M., Franconeri, S., Theeuwes, J., & Suzuki, S. (2010). Characteristic sounds make you look at target objects more quickly. *Attention, Perception, & Psychophysics*, *72*, 1736–1741. doi: 10.3758/APP.72.7.1736

Kiss, M., & Eimer, M. (2011). Attentional capture by size singletons is determined by top-down search goals. *Psychophysiology*, *48*, 784–787. doi: 10.1111/j.1469-8986.2010.01145.x

Kiss, M., Grubert, A., & Eimer, M. (2013). Top-down task sets for combined features: Behavioural and electrophysiological evidence for two stages in attentional object selection. *Attention, Perception, & Psychophysics*, *75*, 216–228.

Kiss, M., Grubert, A., Petersen, A., & Eimer, M. (2012). Attentional capture by salient distractors during visual search is determined by temporal task demands. *Journal of Cognitive Neuroscience*, *24*, 749–59. doi: 10.1162/jocn_a_00127

Lamy, D., Leber, A., & Egeth, H. E. (2004). Effects of task relevance and stimulus-driven salience in feature-search mode. *Journal of Experimental Psychology: Human Perception & Performance*, *30*, 1019–1031. doi: 10.1037/0096-1523.30.6.1019

Leblanc, E., Prime, D. J., & Jolicoeur, P. (2008). Tracking the location of visuospatial attention in a contingent capture paradigm. *Journal of Cognitive Neuroscience*, *20*, 657–671. doi: 10.1162/jocn.2008.20051

Lien, M.-C., Ruthruff, E., Goodin, Z., & Remington, R. W. (2008). Contingent attentional capture by top-down control settings: Converging evidence from event-related potentials. *Journal of Experimental Psychology: Human Perception & Performance*, *34*, 509 –530. doi: 10.1037/0096-1523.34.3.509

Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception & Performance*, *20*, 1000–1014. doi: 10.1037/0096-1523.20.5.1000

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281. doi: 10.1038/36846

Matusz, P. J., & Eimer, M. (2011). Multisensory enhancement of attentional capture in visual search. *Psychonomic Bulletin & Review*, *18*, 904–909. doi: 10.3758/s13423-011-0131-8

Mazza, V., Turatto, M., Umiltà, C., & Eimer, M. (2007). Attentional selection and identification of visual objects are reflected by distinct electrophysiological responses. *Experimental Brain Research*, *181*, 531–536. doi: 10.1007/s00221-007-1002-4

McDonald, J. J., Teder-Sälejärvi, W. A., Di Russo, F., & Hillyard, S. A. (2003). Neural substrates of perceptual enhancement by cross-modal spatial attention. *Journal of Cognitive Neuroscience*, *15*, 10–19. doi: 10.1162/089892903321107783

Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, *35*, 99–115. doi: 10.1111/1469-8986.3510099

Olivers, C. N. L., & Eimer, M. (2011). On the difference between working memory and attentional set. *Neuropsychologia*, *49*, 1553–1558. doi: 10.1016/j.neuropsychologia.2010.11.033

Seiss, E., Kiss, M., & Eimer, M. (2009). Does focused endogenous attention prevent attentional capture in pop-out visual search? *Psychophysiology*, *46*, 703–717. doi: 10.1111/j.1469-8986.2009.00827.x

Shulman, G. L., D'Avossa, G., Tansy, A. P., & Corbetta, M. (2002). Two attentional processes in the parietal lobe. *Cerebral Cortex*, *12*, 1124–1131. doi: 10.1093/cercor/12.11.1124

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experiment Psychology: Human Perception & Performance*, *22*, 1005–1030. doi: 10.1037/0096-1523.22.4.1005

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*, 255–267. doi: 10.1038/nrn2331

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136. doi: 10.1016/0010-0285(80)90005-5

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. doi: 10.1037/0096-1523.34.5.1053

Werner, S., & Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *The Journal of Neuroscience*, *30*, 2662–2675. doi: 10.1523/JNEUROSCI.5091-09.2010

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238. doi: 10.3758/BF03200774

Wolfe, J. M. (2007). Guided Search 4.0: Current progress with a model of visual search. In W. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–120). New York, NY: Oxford.

Woodman, G. F., Arita, J. T., & Luck, S. J. (2009). A cuing study of the N2pc component: An index of attentional deployment to objects rather than spatial locations. *Vision Research*, *1297*, 101–111. doi: 10.1016/j.brainres.2009.08.011

(Received February 18, 2013; Accepted May 15, 2013)