Atten Percept Psychophys (2015) 77:2356–2376
DOI 10.3758/s13414-015-0920-7

Cross-modal cueing in audiovisual spatial attention

Steven P. Blurton[1,2] · **Mark W. Greenlee**[1] · **Matthias Gondan**[2]

Published online: 22 May 2015

Abstract Visual processing is most effective at the location of our attentional focus. It has long been known that various spatial cues can direct visuospatial attention and influence the detection of auditory targets. Cross-modal cueing, however, seems to depend on the type of visual cue: facilitation effects have been reported for endogenous visual cues while exogenous cues seem to be mostly ineffective. In three experiments, we investigated cueing effects on the processing of audiovisual signals. In Experiment 1, we used endogenous cues to investigate their effect on the detection of auditory, visual, and audiovisual targets presented with onset asynchrony. Consistent cueing effects were found in all target conditions. In Experiment 2, we used exogenous cues and found cueing effects only for visual target detection, but not auditory target detection. In Experiment 3, we used predictive exogenous cues to examine the possibility that cue-target contingencies were responsible for the difference between Experiment 1 and 2. In all experiments, we investigated whether a response time model can explain the data and tested whether the observed cueing effects were modality-dependent. The results observed with endogenous cues imply that the perception of multisensory signals is modulated by a single, supramodal system operating in a top-down manner (Experiment 1). In contrast, bottom-up control of attention, as observed in the exogenous cueing task of Experiment 2, mainly exerts its influence through modality-specific subsystems. Experiment 3 showed that this striking difference does not depend on contingencies between cue and target.

Keywords Multisensory processes · Math modeling · Attention · Space-based

✉ Steven P. Blurton
steven.blurton@psy.ku.dk

[1] Institute for Psychology, University of Regensburg, Regensburg, Germany

[2] Department of Psychology, University of Copenhagen, Øster Farimagsgade 2A, 1353 København K, Denmark

Introduction

Everyday perception commonly involves several senses (Welch & Warren, 1986). Multisensory research deals with their interplay, how information obtained by the different sensory systems is integrated, and how attention can be directed to an object of interest, irrespective of whether we see or hear it. Attention can be directed overtly (e.g., by eye, head, or body movements) or covertly, that is, without directing the eyes to the object of interest. Covert orienting of visuospatial attention has been investigated in a large number of studies with spatial cues (Posner, Nissen & Ogden, 1978; Posner, Snyder & Davidson, 1980; Jonides, 1981; Theeuwes, 1991; Carrasco, Ling & Read, 2004; Talsma, Senkowsi, Soto-Faraco & Woldorff, 2010).

Two main classes of cues and their effect on attention are distinguished in the research on covert shifts of attention: endogenous and exogenous cues (Jonides, 1981; Theeuwes, 1991). The properties of the cues play a key role in the modulation of attention. An *endogenous cue* is typically a central, symbolic cue (e.g., an arrowhead); it is supposed to influence attention in a top-down manner. The cued location has to be derived from the arrow, and attention has to be directed deliberately to the cued location. For effective control of voluntary attention, the cue has to be predictive; that is, the (posterior) probability that a target appears at the cued location must be greater than the (prior) probability that a target appears at that location anyway.

Besides this voluntary control of attention, there is a simpler, automated form of attentional modulation. If a cue suddenly appears in the periphery, it automatically attracts visuospatial attention (*exogenous cues*). This is a form of bottom-up control of attention; effective cues are therefore, in general, nonsymbolic and are presented in the periphery close to the possible location of a subsequent target. Such peripheral cues do not even have to be predictive; exogenous cues can direct attention without any validity (Yantis & Jonides, 1984). Both types of cues can effectively draw attention to the cued location. If a subsequent target is presented at that location, responses are faster and more accurate than responses to targets at uncued locations. Such cueing effects are typically interpreted as attention effects: It is assumed that spatial attention facilitates the perception of stimuli presented at the cued location. While some studies focused on response speed (Posner, 1980; Posner, Snyder & Davidson, 1980; Jonides, 1981), other studies used perceptual decision tasks to test response accuracy (Bashinski & Bacharach, 1980; Dosher & Lu, 2000; Carrasco, Ling & Read, 2004; Liu, Fuller & Carrasco, 2006; Fairhall & Macaluso, 2009).

The initial studies on the spatial cueing effect focused on the visual modality (but see Lansman, Farr & Hunt, 1984). Subsequent research also investigated the effects of secondary tasks, such as language processing (Posner, Inhoff, Friedrich, & Cohen 1987) on detection times in the spatial cueing task. Multisensory spatial attention was further investigated to identify the extent of links between modality-specific attentional systems and whether there is a supramodal attentional system (Farah, Wong, Monheit & Morrow, 1989; Eimer & Schröger 1998; Eimer, 2001; Eimer & Driver, 2001; Talsma & Kok, 2002; Talsma et al., 2010).

Using a perceptual discrimination task, Spence and Driver (1996) investigated the effects of endogenous visual cues on visual and auditory perception. Critically, they used an orthogonal cueing technique (Spence & Driver, 1994); an improved method for cueing in perceptual decision tasks that precludes response priming (Ward, 1994; Spence & Driver, 1997). Spence and Driver (1996) reported cueing effects on both visual and auditory targets, that is, visual cues directed cross-modal attention and thus also speeded up the perceptual decision in pure auditory targets. Response accuracy was less affected by the attentional manipulation in either modality (Spence & Driver, 1996; Table 1). Spence and Driver (1996) argued that auditory perception is affected by attentional mechanisms, but only through late processes that do contribute to perceptual decisions but not to simple reaction time. Several EEG studies, however, found that very early processes were equally affected by attentional manipulations (Eimer, 2001; Talsma & Woldorff, 2005).

The effects of cross-modal cueing in exogenous attention were different: For example, Buchtel and Butter (1988) used visual and auditory exogenous cues and measured the response time for visual and auditory targets. The main finding of their study was that visual—but not auditory—targets were affected by the preceding visual cues. In line with this, Spence and Driver (1997) conducted a series of experiments in which they demonstrated that exogenous visual cues did not influence the perception of auditory signals. The relationship between audition and vision in exogenous spatial attention is asymmetric—that is, exogenous auditory cues influence visual perception (Spence & Driver, 1994) but visual cues do not seem to affect auditory perception (Driver & Spence, 1998). These results are challenged by those of Ward (1994) and Ward, McDonald, and Lin (2000) who reported exactly the opposite asymmetry, albeit in different experimental setup with rather complex cueing and a nonspatial go/no-go task. The nonspatial response task was subsequently identified as the important difference between those studies (Spence, McDonald & Driver, 2004; Koelewijn, Bronkhorst & Theeuwes, 2010).

Although these studies investigated cross-modal cueing effects, they provide little insight into the interplay between the processing of multisensory information and attention shifts—unless one takes up the position that multisensory integration *is* cross-modal attention (Spence, McDonald & Driver, 2004, p. 306ff). Like in spatial cueing experiments, it has been demonstrated that exogenous and endogenous multisensory cues have quite different effects on multisensory integration (for a review, see Talsma et al., 2010). Endogenous cues seem to facilitate multisensory integration of subsequent percepts (Fairhall & Macaluso, 2009), especially if the perceptual load is high (Talsma & Woldorff, 2005). Studies with exogenous cues, on the other hand, support the notion that these cues induce a location unspecific spread of spatial attention across modalities (Talsma et al., 2010). For example, a spatially uninformative auditory cue can enhance target detection in a visual search task, whereas temporal, but spatially uninformative, visual cues do not seem to have any effect on performance in visual search (Van der Burg et al. 2008). It remains open whether this cross-modal spread of attention is symmetric, that is, if visual cues without spatial information can facilitate performance in an auditory task (e.g., discrimination task). The series of experiments by Spence and Driver (1997) would suggest that this is not generally the case.

A common experimental paradigm to study multisensory perception is the redundant signals setup. In this setup, participants are provided with information in different modalities, for example, vision and audition. Participants are usually given the task to respond as quickly as possible to any stimulus (Miller, 1982; Diederich & Colonius, 1987). The typical finding is that if signals from both modalities are present (redundant signals, AV), responses are faster than to targets from any single modality (unimodal targets), that is, faster than responses to auditory (A) or visual (V) targets. A more precise analysis of the mechanisms involved in the integration of redundant signals benefits from the extension of this basic setup

Table 1 Experiment 1—Results of signal detection theory parameter estimation and frequency of false alarms

	Participant															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
β	**0.58**	**0.71**	1.59	0.87	1.09	1.34	**0.36**	1.07	1.22	2.70	1.53	2.11	2.20	1.56	1.59	1.77
d'_{valid}	**3.54**	**3.21**	∞	∞	3.92	3.94	**3.53**	4.03	∞	∞	∞	∞	∞	∞	4.76	∞
$d'_{invalid}$	**3.18**	∞	∞	3.71	∞	4.17	**3.20**	3.67	4.05	5.53	∞	∞	∞	∞	4.42	4.60
FA (%)	**28.1**	**24.0**	5.6	19.1	13.9	0.9	**35.8**	14.2	11.1	0.4	6.3	1.7	1.4	6.0	5.6	3.8

β, response criterion set by the participant: lower values represent more liberal criteria; d', sensory sensitivity: higher values represent higher sensitivity, infinite values mean perfect discriminability; FA, false-alarm rate (% of responses to catch-trials)

to presentation with onset asynchrony (e.g., V100A, with the visual target component preceding the auditory target component by 100 ms; Miller, 1986). The unimodal conditions A and V can then be considered as end points of a continuum where the second stimulus follows with infinite SOA. Whereas attentional effects on visual perception have been investigated with redundant signals (Mordkoff & Yantis, 1993; Feintuch & Cohen, 2002; Miller, Beutinger & Ulrich, 2009; Mordkoff & Danek, 2011), the effects of spatial cues on the perception and integration of multisensory redundant signals are largely unknown. Only one study investigated differences in multisensory processing of redundant signals under different spatial attention conditions (Gondan, Blurton, Hughes & Greenlee, 2011). In that study, the effect of a different attentional focus was compared in two conditions (wide vs. narrow focus). As one would expect, a narrow focus facilitated the detection of audiovisual signals and their integration into a single percept. However, the study did not explore the nature of cross-modal interactions in spatial attention.

In the present study, we investigated the effect of visual spatial cues on multisensory perception of auditory, visual, and audiovisual targets at different peripheral locations. In Experiment 1, we used central, symbolic, and informative (endogenous) visual cues. In Experiment 2, we employed peripheral and noninformative (exogenous) visual cues. In Experiment 3, we also used exogenous cues; however, these were again informative. A diffusion superposition model (Schwarz, 1994; Diederich, 1995) was fitted to the observed response times to assess the effects of those cues on multisensory integration of redundant signals. Using comparisons of hierarchically nested models, we tested for the presence of cueing effects and, if present, whether these effects were modality-invariant or modality-specific.

Experiment 1

In Experiment 1, participants made speeded responses in an audiovisual detection task with centrally presented predictive visual cues. In addition to unimodal auditory and visual target conditions, we employed audiovisual target conditions in which the two stimulus components were presented with onset asynchrony.

Methods

Participants

Sixteen students (mean age: 23.9 [range 20–30] years; 4 males, and 1 left-handed) of the University of Regensburg participated in Experiment 1. All had normal or corrected-to-normal visual acuity and reported normal hearing. All were naïve with regards to the purpose of the experiment. They were either paid for participation (7 € per hour) or received course credit. Before participation, they gave written, informed consent. The study was conducted in accordance with the Declaration of Helsinki.

Experimental task

Participants were instructed to respond by pressing a response button as fast as possible when an auditory and/or visual target appeared either at the expected (cued) location or at the unexpected (uncued) location and to withhold responses in catch trials with only cues but no targets (simple speeded response). They were instructed to fixate the central position of the screen, which was marked by a cross during the whole experiment. On each trial (targets and catch trials), participants had 1 second to respond before the trial ended and stimulation resumed with the next trial.

Apparatus and stimuli

The experiment took place in a light and sound attenuated room that was dimly illuminated from behind and above (Industrial Acoustics GmbH, Niederkrüchten, Germany). The participants viewed the screen with their head resting on a chin rest 70 cm in front of the screen that had a size of 57 cm × 72 cm (54.4 deg) and was back-projected onto by a projector (NEC V230X, NEC Corporation, Minato, Japan) from outside

the cabin. The visual target was a Gabor patch (sigma = 0.8 deg, 1.8 cycles/deg, Michelson contrast: L_M = 0.969, size: 6 deg) presented either in the left or right periphery on the horizontal meridian. To discourage eye movements, we used a relative high eccentricity (24 deg). Visual targets were presented on a uniform gray background with a white fixation cross (0.8 deg) at the center of the screen. The auditory target was white noise (45 dBA) with a short (5 ms) ramp-on (and off) to avoid clicking noises at onset and offset. The loudspeakers were mounted adjacent to the screen at the left and right edges of the screen (34 deg eccentricity). The speakers were connected to a low-latency sound card (Soundblaster Audigy 2 ZS, Creative Technology, Singapore), which was installed in a standard IBM-compatible PC running Presentation (Neurobehavioral Systems, Berkeley, CA). The target stimuli were generated with MATL AB (MathWorks, Natik, MA).

The redundant targets were presented with stimulus onset asynchrony (SOA). The auditory target preceded the visual target by 33 ms, 67 ms, or 100 ms, or followed the visual target by the same amount of time. Thus, redundant targets were presented with seven SOAs: A33V (i.e., auditory target followed by visual target with 33 ms SOA), A67V, A100V, V33A, V67A, V100A, and the synchronous target condition, AV. Together with the unimodal conditions A and V ("infinite" SOA), we had nine SOA conditions.

The spatial cues were arrowheads (1 deg) presented adjacent to the central fixation cross. The arrows were presented for 67 ms and directed either to the left side or to the right side. The cueing interval, that is, the time between cue offset and target onset was 300 ms. Each trial started with a cue followed by a single target (unimodal visual or auditory condition), by an audiovisual target (redundant visual-auditory condition), or no target (catch trials). Both the visual and the auditory targets were presented for 233 ms (no masking). To discourage participants from executing anticipatory responses, we randomly presented one catch trial in six target trials. In the catch trials, no targets were presented and the stimulation program waited for 1,000 ms before proceeding with the next trial. In target trials, targets were presented at the cued location in 75% of the trials (cue validity). Between two trials we used a variable inter-trial-interval that had a fixed base time of 600 ms, plus an exponentially distributed random duration (expected time: λ^{-1} = 1000 ms). The ITIs were generated by a uniform random generator $\mathbf{U} \in (0, 0.95)$ (the upper boundary was chosen to avoid overly long intervals) and then transformed to the inter-trial interval by $-\ln(\mathbf{U}) / \lambda$. Participants should press a response button placed under the index finger of their dominant hand.

Procedure

Each participant was tested on 3 days in separate sessions; each session lasted approximately 45 min. In each session, participants were tested with two blocks of 336 trials each. A short break was given between the blocks. For each participant, we obtained data from 48 trials with an invalid cue and from 144 trials with a valid cue. These numbers are pooled for the left and right target locations but apply to all SOA conditions employed. Thus, we tested each participant in 1,728 cue-target trials and 288 catch trials.

Analysis of response times

The high number of catch trials allows for a detailed analysis of anticipations in the empirical response time (RT) distributions. First, we estimated parameters of the signal detection model (Green & Swets, 1966) by the amount of responses to catch trials ("false alarm") and correct target detection ("hits"). The goal is to provide an overview of the participants' strategies; three participants with very low response criteria ($\beta < 0.8$)—indicative of an overly guessing strategy—were excluded from further analysis.

For the main analysis, we used the so-called "kill-the-twin procedure" (Eriksen, 1988) combined with the Kaplan-Meier estimate to retrieve the RT distribution swept from anticipations (Koch et al., 2013). The reasoning behind this procedure is that the real empirical distributions of target RTs are contaminated by guesses and anticipatory responses. Some of the guesses occur in catch trials; in this case, they are easily detected. A number of guesses occur in targets, although these can be considered "lucky" guesses. These "lucky" guesses lead to an interruption of the response process, which otherwise would have continued (i.e., it is a censored observation). Assuming stochastically independent guesses and informed responses, an unbiased estimate of the hidden RT is obtained by the Kaplan-Meier method known from survival analysis (Kalbfleisch & Prentice, 2002). The assumption of an uninformative censoring mechanism yields a conservative bound for the correction of the response time distribution (Gondan & Fimm 2013). To control for contaminant guesses, we applied the kill-the-twin procedure and censored six correct responses in closest temporal correspondence to a given guess in a catch trial, thereby accounting for the fact that for each catch trial 4.5 valid and 1.5 invalid trials were presented, respectively. In an analogous manner, we also censored all RTs greater than 1,000 ms (<0.5% of RT in all participants) to correct for attentional lapses ("misses"). No further correction was applied to the RT data. After pooling responses from left and right targets, the Kaplan-Meier estimate $\hat{G}(t)$was used to estimate the distribution of informed response times; the area under $1 - \hat{G}(t)$ was then used to calculate mean RT and SD.

The resulting mean RTs were entered into a 2 × 9 (Cue Validity × SOA) repeated measures ANOVA. This ANOVA was used to assess cueing and modality effects as well as their interaction. Post-hoc two-sided paired *t* tests were conducted for pairwise comparisons (FDR corrected for

multiple comparisons, Benjamini & Hochberg, 1995). We tested all SOA conditions for cueing effects, but for the sake of parsimony, effects of SOA were tested only for the unimodal conditions (A vs. V). For the remaining SOA conditions, we refer to the main analysis of the Diffusion Superposition Model (DSM, Schwarz, 1994), which was then fitted to the mean reaction times in all conditions.

Race-model inequality

Several models have been proposed to account for the redundant signals effect; the most prominent model classes are the race models (Raab, 1962) and coactivation models (Miller, 1982). In the former, it is assumed that detection of signals can be conceived as a race between active channels in which the winner of the race determines the detection time. The redundant signals effect is then a consequence of *statistical facilitation*: if the latency distributions of detection times of the two channels overlap, detection of redundant signals is, on average, faster, because in that case slow detection times in one channel can be compensated for by faster detection in the other channel. It has been shown that, under certain assumptions ("context invariance"; e.g., Luce, 1986, p. 129), the redundancy gain by the race model has an upper bound (Miller, 1982, 1986). Because information is never integrated in the sense that it is pooled into a common channel, the response time distribution function of redundant target, $G_{AV}(t)$, cannot exceed the sum of the distribution functions of the unimodal targets:

$$G_{AV}(t) \leq G_A(t) + G_V(t), \text{ for all } t. \quad (1)$$

Violations of this upper bound often have been reported and are usually interpreted as evidence for coactivation (Miller, 1982, 1986; Schröter, Ulrich & Miller, 2007). We tested the race-model inequality with the permutation test described by Gondan (2010). For the permutation test, we used an aggregated test statistic, which was the weighted sum $\sum_\tau w_\tau \Delta_\tau$ of SOA-specific violations Δ_τ of the race-model inequality (Gondan, 2009), for example $\Delta_{33} = \max[0, \hat{G}_{A33V}(t) - \hat{G}_A(t) - \hat{G}_V(t - 33)]$ for condition A33V. The weighting function had the form of a shifted umbrella; specifically, we assigned weights $w_\tau = 2$, 3, 4, 3, 2, 1, and 1 to conditions V100A, V67A, V33A, AV, A33V, A67V, and A100V, respectively. The weight function was shifted, because from the difference in mean reaction time of single target conditions we expected race-model violations to be most pronounced when the visual stimulus preceded the auditory stimulus by moderate SOA ("psychological synchrony," Hershenson, 1962)—that is, V33A. For each condition, we calculated violation statistics Δ_τ for six RT percentiles (5th, 10th, …, 30th percentile) and standard one-sample t statistics for each percentile aggregated over participants. To avoid multiple tests, the six t statistics were aggregated into a T_{max} statistic using the maximum for the six percentiles. The distribution of this T_{max} statistic under the null hypothesis (i.e., the race model holds) was retrieved by a permutation procedure (Gondan, 2010). The p values were then calculated as the proportion of permutations that yielded T^*_{max} greater than the observed T_{max}. The race model was significantly violated, if the p value was less than 5%—that is, if the observed T_{max} was greater than 95% of the simulated T^*_{max}.

Diffusion superposition model

We used the diffusion superposition model (DSM, Schwarz, 1994) to assess both redundancy gains and attention effects on response times. This computational model describes response times of a redundant signals experiment and can be applied readily to data from two or more experimental manipulations, such as spatial cueing. The DSM assumes accumulation of information over time that can be described by a Wiener process with drift $\mu > 0$ and diffusion coefficient $\sigma > 0$ in the presence of an absorbing barrier c (i.e., the response criterion). In a redundant signals task, each sensory processing channel is assumed to represent such a process, that is, a Wiener process with parameters μ_A and σ_A for the auditory processing channel and parameters μ_V and σ_V for the visual processing channel. The detection time **D** is the first passage time to c, it follows an inverse Gaussian distribution with expected value $E(\mathbf{D}) = c/\mu$ (Cox & Miller, 1965, p.222). When redundant (audiovisual) targets are presented, both channels are active and it is assumed that the information of both channels is pooled into a common channel (additive superposition). The DSM is thus an instance of a coactivation model; the redundant signals effect is explained to be due to the faster buildup of evidence in the case of both channels being active. The contributions of both processes are additively superimposed, so that the drift of the common process is $\mu_{AV} = \mu_A + \mu_V$ (Fig. 1, solid lines).

Because the diffusion process is supposed to describe only the detection time and response time is usually assumed to be a compound of an information accumulation part and residual process ("motor processes", Luce, 1986), the mean latency of all those processes (μ_M) was added to the above expression to derive predictions for the observed mean response times. The response criterion c was fixed at 100, because it is only a scaling factor.

The model fit was assessed by a X^2 statistic that was the sum of normalized differences between observed and predicted mean response times. This statistic is asymptotically χ^2 distributed with degrees of freedom (df) determined by the number of predicted experimental conditions minus the number of free parameters. Of main interest in this study were three models that differed only with respect to the assumed spatial cueing effects. The first and most restrictive model had five parameters: four diffusion parameters (μ_A, σ_A, μ_V, σ_V) and the mean latency of residual processes μ_M. This model served as a null-model, because it contained no free parameter

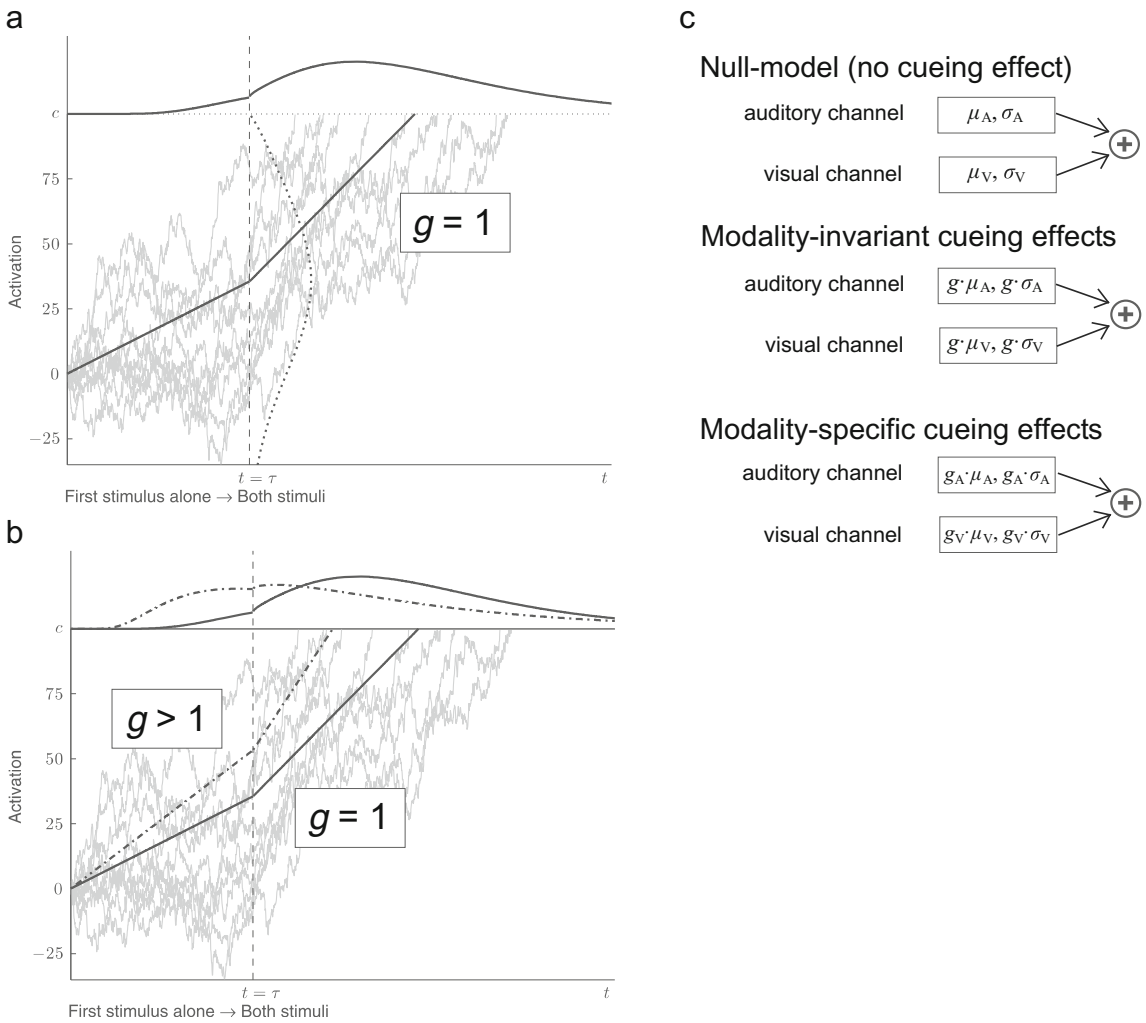

Fig. 1 **a**: The signal detection process in the common channel according to the Diffusion Superposition Model. At stimulus onset, a Wiener process with drift starts at X(0) = 0 and evolves over time, until the criterion c is hit for the first time. Displayed are ten realizations of a Wiener processes with parameters $\mu_1 = 0.53$ and $\sigma_1 = 4.3$ and response criterion $c = 100$. When redundant targets are presented with some SOA, the process contains two parts: at the beginning, only one channel contributes to the activity of the common channel. At the onset of the second stimulus component, the process has attained a state $\mathbf{X}(t=\tau) < c$, if the criterion has not already been reached (vertical distribution). Afterwards, the second channel ($\mu_2 = 0.53$) also is active and its contribution is added into the common channel. The effects of the additive superposition can be seen in the average position of a process (i.e., the solid linear trend) and in the solid upper curve that represents the resulting first-passage time density that is arbitrarily scaled for display purposes. **b**: The effect of the attention scaling factor g in the DSM. The solid linear trend and the solid curve are the same as in a. The dashed figure parts show the effect of an attention scaling factor g (here: $g = 1.5$) with other parameters kept constant. It is easily seen that the criterion is reached earlier, on average. **c**: Schematic overview of the three specified and tested models. The models differ only with respect to the attention scaling factor g for the detection of targets presented at pre-cued locations. In the null-model, the cueing factor is constant ($g = 1$). The cueing factor is variable, but the same across sensory modalities in the modality-invariant attention model, and is variable and possibly different across the sensory modalities in the modality-specific attention model

for cueing effects and, thus, did not predict any effects of attention shifts. The second and more plausible model had one additional free parameter, a separate attention factor g for the processing of validly cued targets. With this additional parameter, cueing effects could be modelled by facilitation in the processing channel(s) by efficient allocation of spatial attention after validly cued trials. If processing of the target is facilitated because it is presented at an expected location, g should be greater than unity and the model predicts faster detection and response times (Fig. 1b, dot-dashed lines). This model was still rather restrictive, because the experimental manipulation of cue validity of all nine target conditions was allowed to affect only this single parameter. Thus, this model captured possible cueing effects as a modality invariant attention effect. To test whether this assumption was justified by the data, we fitted a third model with an additional free parameter to allow for modality-specific cueing effects. In each experiment, we fitted the models to mean response times observed in all 18 experimental conditions. For clarity, we present aggregate model fits of all participants.

In Experiment 1, the aggregate model fit of the null-model with five free parameters contained $N \times 18 = 234$ mean response times of $N = 13$ participants (3 participants were excluded, see below) from which $N \times 5 = 65$ free parameters were estimated, resulting in $df = 234 - 65 = 169$ degrees of freedom (df). The aggregate model fit of the modality-invariant cueing-effect model had thus $df = 156$, that of the modality-specific cueing effects model $df = 143$. The influence of the additional parameters were judged to be significant on group level, if the observed difference ΔX^2 was > 95% percentile of the $\chi^2(13)$ distribution for each parameter added. Details about the nested model tests can be found in the Appendix.

Results

Analysis of response times and anticipations

Responses to catch trials (false alarms) were quite frequent. Most participants responded on 0.4% up to 20% of the catch trials; however, three participants responded on more than 25% of catch trials. The sensory discriminability measures and evidence criterion estimates of SDT are given in Table 1. As one would expect, discriminability was rather high (> 3.0 in all participants); however, the three mentioned participants chose the most liberal evidence criteria β of all participants. Censoring was employed to counteract anticipatory effects; however, these participants (Table 1) had to be removed from further analyses, because they exhibited such a high false-alarm rate, which would have required us to censor a large amount of the RT distribution. All subsequent results and model fits therefore are based on 13 participants.

The main effects of Cue Validity ($F(1,12) = 43.13$; $p < 0.001$) and SOA ($F(1,12) = 125.5$; $p < 0.001$) were significant, as well as the interaction between Cue Validity and SOA ($F(1,12) = 13.37$; $p = 0.003$). As expected, responses in all target conditions were faster, on average, if the target was preceded by a valid spatial cue than by an invalid cue. The cueing effect was significant in the unimodal auditory and visual target conditions as well as in all audiovisual targets (Table 2). Validly cued auditory targets ($M = 227 \pm SD = 20$ ms) were not only faster than invalidly cued auditory targets (246 ± 29 ms), but also significantly faster ($t(12) = 7.97$, $p < 0.001$) than validly cued visual targets (269 ± 29 ms; modality effect, Fig. 2a). Invalidly cued auditory (246 ± 29 ms) and visual targets (324 ± 30 ms) also showed an effect of target modality ($t(12) = 14.48$, $p < 0.001$). Even though both modality-specific cueing effects were significant, there was a numerically greater cueing effect in visual targets than auditory targets (Fig. 2a). Regarding the redundant targets, we found the typical wing-shaped patterns of mean RT in both the valid cue condition and the invalid cue condition (Fig. 2b). Within both cue validity conditions, responses were fastest in the (synchronous) redundant target condition (Table 2). With increasing delay of either target component, mean response times approached the respective mean response time of the unimodal targets.

Race-model inequality

The race-model inequality was significantly violated on group level in the invalid cue condition ($T_{max} = 2.60$; critical $T^*_{max} = 2.38$; $p = 0.032$) but not in the valid cue condition ($T_{max} = 0.543$; $T^*_{max} = 2.41$; $p = 0.523$). Only in the invalid

Table 2 Experiment 1—Mean (*MRT*, ms) and standard deviations (*SD*) of response times for auditory (A), visual (V), and audiovisual (AV) targets presented with onset asynchrony after valid cues and invalid cues

Target condition	Cue condition				Cue validity effect: $MRT_{invalid} - MRT_{valid}$	
	valid		invalid			
	MRT	*SD*	*MRT*	*SD*	*t*(12)	*p*
A	227	20	246	29	2.588	0.024*
A100V	219	17	241	25	5.846	<0.001*
A67V	216	15	233	19	4.485	<0.001*
A33V	211	13	235	21	10.482	<0.001*
AV	206	12	225	23	3.670	0.003*
V33A	229	10	248	21	4.831	<0.001*
V67A	244	12	270	19	7.088	<0.001*
V100A	255	14	290	22	5.835	<0.001*
V	269	29	324	30	7.845	<0.001*

Results are based on data from 13 participants. Participant-specific mean RTs were retrieved from RT distributions corrected for anticipations (see *Methods*). For valid cues, these distributions included 144 responses per participant; for invalid cues 48 responses (75% cue validity). The last two columns are results from two-sided paired *t* tests for a cueing effect ($MRT_{valid} - MRT_{invalid}$) in the respective SOA condition. Significant differences (after FDR-correction) are marked with an asterisk. The *MRT* data corresponds to the data points depicted in Fig. 2a, b

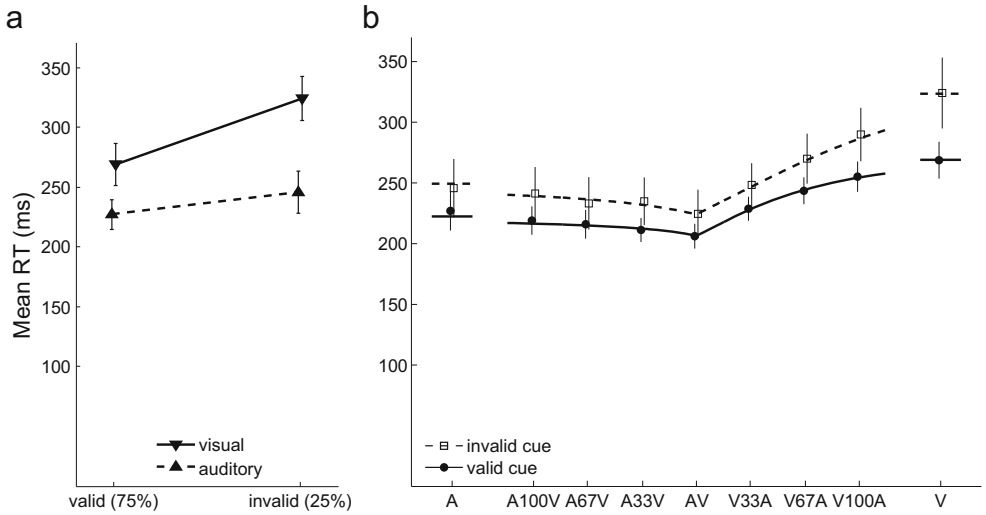

Fig. 2 Experiment 1—**a**: Mean response time (± 95% confidence intervals) of auditory and visual targets following valid and invalid cues. Both modality effects and cueing effects were significantly different from zero. **b**: Mean response times observed in valid (circles) and invalid cue trials (squares) averaged across all participants with error bars denoting average of estimates of the 95% confidence intervals. The lines represent model predictions of the DSM with a modality-invariant cueing effect for the valid (solid line) and invalid (dashed line) cueing condition. The model predictions also are averaged across all participants

cue conditions redundancy gains were significantly greater than predicted by the race model. In the valid cue condition, redundancy gains were small in size and compatible with the race-model assumption.

Diffusion superposition model

All response conditions were used to fit the diffusion superposition model to the observed mean response times. Different model assumptions were made to account for different cueing effects. The null-model, assuming no effects of spatial cueing, was clearly rejected (X^2 = 1481; df = 169; $p < 0.001$). In contrast, both the model with modality-independent cueing effects (X^2 = 169.2; df = 156; p = 0.222; Fig. 2b) and the model with modality-specific cueing effects (X^2 = 157.0; df = 143; p = 0.200) both provided good fits to the observed data. In line with the ANOVA results, the comparison of the null-model with the modality-invariant cueing effect model (Fig. 1c) yielded a significant difference (ΔX^2 = 1311; df = 13; $p < 0.001$), indicative of a cueing effect: The speed of processing of targets at expected locations was increased and both models with an attention scaling factor g yielded very good predictions for the observed data (Table 3). The model predictions follow the same wing-shaped pattern as observed in the data and, more importantly, the models predict a decrease of response times for targets presented at expected locations because of the faster processing for those targets than at unexpected locations ($g > 1$).

The additional parameter for modality-specific attention effects (Fig. 1c) improved the fit only marginally (ΔX^2 = 12.26; df = 13; p = 0.506). In other words, the attention scaling factors g_A and g_V were not significantly different from each other. The same holds true for the mean latency of all nonperceptual processes (ΔX^2 = 17.39; df = 13; p = 0.182). Thus, nonperceptual processes were not affected by cue validity, at least to a much lesser extent than perceptual processes. The modality-invariant model predicted the observed patterns of response times very well with the attention scaling factors fixed to be equal across modalities (Fig. 2b). Consistent with faster responses to auditory targets, drift and variance of the auditory processing channel were somewhat greater than drift and variance of the visual processing channel (Table 3), which largely resembled results that were reported earlier in simple-response divided attention tasks (Schwarz, 1994).

Table 3 Experiment 1—Parameter estimates and model fit of the Diffusion Superposition Model with modality invariant cueing effect

Parameter	Average	SD
μ_A	1.60	0.81
σ_A	26.5	25.8
μ_V	0.71	0.17
σ_V	3.6	2.0
g_A=g_V	1.72	0.61
μ_M	175.5	15.7
Goodness of fit (*higher X^2 indicate worse fit*)		
X^2(156)	169.2	p = 0.222
Comparison with modality-specific cueing effects ($g_A \neq g_V$)		
ΔX^2(13)	12.26	p = 0.506

μ_A, σ_A, μ_V, σ_V: drift and diffusion constant for auditory and visual diffusion processes; g: attention factor (facilitation) for validly cued targets; μ_M: mean latency of residual processes

Discussion

In Experiment 1, we extended the cross-modal spatial attention setup (Spence & Driver, 1996) to a redundant signals setup with visual spatial cues. This provided a more detailed picture of the effects of those types of cues on multisensory perception, because one cannot only investigate the presence or absence of redundancy gains, but rather the mechanism of multisensory integration with fine-grained temporal resolution (Miller, 1986). The motivation for this extension was twofold: first, on the empirical side, we obtained information about multimodal targets and, on this note, unimodal targets can be viewed as extremes of the continuum of bimodal targets (namely, with SOA = ± ∞). The redundant signals setup is then a natural extension of previous experimental designs investigating cross-modal attention. In this way, we also could test race-model predictions and found significant violations of the race-model inequality, however, only for invalidly cued targets. The redundancy gain observed in the valid cue condition was smaller and did not significantly violate race-model predictions. Besides general issues of power in testing the race-model inequality at multiple percentiles, we do not have a straightforward explanation of this discrepancy (Kiesel, Miller & Ulrich, 2007). Simulations using the parameter estimates of Table 3 showed indeed that for the sample size of the present experiment the amount of race-model violations was rather small (approximately 50%). The facilitation effects observed in the valid cue condition could have led to such increased processing speed that the second stimulus component itself could not substantially speed up the detection processes: The more efficiently a signal is processed the smaller is the expected redundancy gain. By analogy, if targets are presented at different eccentricities, redundancy gains are often reported to be smaller for targets presented at more central locations (i.e., if the target falls into the small receptive fields of the fovea) than for targets presented at more peripheral locations (Schwarz, 2006). In line with this, the pip-and-pop effect (Van der Burg et al., 2008) also has been found to be greater in invalidly cued search displays than in correctly cued ones (Zou, Müller & Shi, 2012).

Sophisticated response time models are available that can explain the redundant-signals effect (Schwarz, 1994; Diederich, 1995). These models provide further insights that cannot be obtained by standard analyses of additive effects on mean RTs (i.e., factorial analysis of variance designs). By application of this model, we demonstrated that endogenous spatial cues not only affected response times to visual and auditory targets but rather to the whole SOA-dependent mean RT curve of multisensory integration (Table 1; Fig. 2b). Regarding the unimodal targets, we replicated findings that were reported earlier. On the one hand, this was the effect of visual cues on visual targets, as described by Posner (1980). On the other hand, we found an effect of cue validity on auditory targets (Spence & Driver, 1996). These results were in line with previous studies (Farah et al., 1989), but by fitting a diffusion model to the data, we revealed that the observed cueing effects could be modeled by a single parameter to describe the effects of attention in both modalities. The validity of this assumption was indicated by the good-to-excellent agreement of model predictions and data. Under the assumption of modality-invariant cueing, the cue was allowed to affect an attentional scaling factor, which in turn affected processing in both modalities to an equal extent. The model with modality-specific attention effects did not provide a better account for the observed data, so, by parsimony, the results from Experiment 1 support the assumption of a modality invariant attention effect.[1]

The attention factor describes the increased efficiency for targets presented at expected locations, that is, when they fall into the spatial attention focus in the valid cue conditions. This leads to faster responses, as less time is needed, on average, to reach the response criterion. Alternatively, the higher efficiency could reflect a constant drift, but a lower response criterion because a lower response criterion is mathematically equivalent to increased drifts and variance of both auditory and visual stimuli, hence the modality-invariance. The interpretation of information sampling efficiency fits well with the notion of attention as a signal-to-noise modulator (Carrasco, 2000; Lu & Dosher, 1998): If the cue pointed to the correct location, participants had enough time to shift their attention to that location so that upon stimulus onset information about the stimulus was extracted more efficiently. Conversely, if the stimulus appeared at the unexpected location, information about that stimulus would have been obtained less efficiently, as indicated by the estimate of the attention factor. Different efficiency also could explain the better performance and increased sensitivity, for example, in detection tasks (Bashinski & Bacharach, 1980; Dosher & Lu, 2000; Carrasco et al., 2004). The interpretation that mainly perceptual processes were facilitated was further supported by the results that indicate that the model could well explain the observed data with constant residual processes latency. The inspection of parameter estimates leads to interesting implications: if endogenous attention is shifted by cues in a way that perceptual processes are modulated independent of modality, this is indicative of a single, a supramodal spatial attention system (Farah et al., 1989). Moreover, if objects appear within this focus, all its features benefit from the increased efficiency, independent of their modality. This is exactly what we observed in Experiment 1. However, preattentive (early) integration of the auditory and visual targets could also explain the observed modality-invariant attention shift. According to this view, multisensory objects are integrated into a single percept at an early

[1] Of course, non-significant results should not be taken as evidence for the absence of an effect, it might also be the case that there were modality-specific effects that were just too weak to reach statistical significance.

stage so that the whole percept benefits from attentional resources. The distinction between attentional facilitation of already integrated multisensory percepts and "true" cross-modal attention (Spence et al, 2004) is not possible by means of the applied response time model. Either way, the results demonstrate that, if stimuli fall into the attention focus, all features benefit from attention in the same manner; both qualitatively and quantitatively.

Experiment 2

In Experiment 2, we studied the effect of exogenous cues on multisensory integration, using the same experimental setup. Unlike Experiment 1, we used peripheral, nonpredictive cues with a shorter cueing interval.

Methods

Participants

Nine new participants (mean age: 29.6 [range 23–37] years; 6 females, and 2 left-handed) were tested in Experiment 2. Before participation, they gave written, informed consent. All were naïve regarding the purpose of the experiment.

Apparatus and stimuli

Experiment 2 was performed with the same apparatus and the same target stimuli as Experiment 1. In Experiment 2, we used red square frames (6 deg) as cues that were presented at the possible target locations. The target followed 100 ms after cue offset. The cueing interval was shorter than in Experiment 1 to avoid the so-called inhibition of return effect (Posner & Cohen, 1984). Contrary to Experiment 1, these cues were noninformative; that is, in approximately 43% (3/7) of the trials the subsequent target was presented at the cued location and with the same frequency at the opposite location. In approximately 14% (1/7) of cued trials, there was no subsequent target (catch trial). The frequency of catch trials was the same as in Experiment 1. Target location also was split evenly between left and right peripheral positions.

Experimental task

Participants were instructed to respond as fast as possible when a target appeared at either location and to avoid anticipations, that is, responding to the cue alone. Again, participants were instructed to maintain their fixation, which was indicated by a cross at the center of the screen.

Procedure

As in Experiment 1, each participant was tested in three separate sessions and each session lasted approximately 45 minutes. The total number of trials for each participant is 1, 728 target trials and 288 catch trials. The number of trials is thus the same as in Experiment 1; however, due to the change in cue validity we obtained 96 replications for each condition, regardless of whether the cue was valid or invalid. In one half of the trials, a cue in the left periphery was displayed; in the other half of trials, a cue appeared in the right periphery.

Statistical analyses

We again employed signal detection theory and the same guessing correction as in Experiment 1, that is, we determined contaminants in the RT distribution by the kill-the-twin procedure and obtained estimates of RT mean and variance with the Kaplan-Meier estimate. The only difference is that due to the changed cue validity we censored three responses for every response to a catch trial for both valid and invalid cue conditions. Attentional lapses (RT < 1,000 ms) also were censored; this affected less than 1% of all RT in all but one participant (Participant 2: 4.2%). The analysis was analogous to Experiment 1: We fitted the diffusion superposition model with different underlying assumptions to the data tested the models fits as before by using a X^2 goodness-of-fit statistic. The permutation test of the race-model inequality was analogous to Experiment 1.

Results

Analysis of response times and anticipations

Observed anticipations—responses to catch trials—were less frequent than in Experiment 1: Participants chose rather conservative response criteria (Table 4). One participant was very liberal in the decision to respond (β = 0.47), so we excluded this participant from further analyses (Table 4). The estimated sensitivity measures *d*' of all participants were again rather high in both the valid and the invalid cue condition.

The ANOVA yielded significant main effects for the two factors: Cue Validity ($F(1,7) = 29.50$, $p = 0.001$) and SOA ($F(1,7) = 47.63$, $p < 0.001$), as well as for the interaction Cue Validity × SOA ($F(1,7) = 9.06$, $p = 0.020$; Fig. 3a). In line with previous studies, we found a cueing effect of visual exogenous cues on the mean response times of visual targets (Table 5; Fig. 3a): mean response time (± *SD*) in the visual target condition decreased from 333 ms (± 36 ms) to 296 ms (± 32 ms), if targets were presented at the cued location rather than the opposite location ($t(7) = 6.715$, $p < 0.001$). However, unlike Experiment 1, there was no cueing effect for the auditory targets ($t(7) = –0.568$, $p = 0.588$): the mean RT of validly cued

Table 4 Experiment 2—Results of signal detection theory parameter estimation and frequency of false alarms

	Participant								
	1	2	**3**	4	5	6	7	8	9
β	0.97	1.09	**0.47**	1.05	0.91	1.38	1.86	1.18	1.18
d'_{valid}	3.43	2.76	**3.72**	3.89	3.75	∞	∞	4.23	∞
$d'_{invalid}$	3.32	2.89	**3.52**	3.66	3.75	4.21	4.91	4.23	4.02
FA (in %)	16.7	13.9	**31.9**	14.6	18.1	8.3	3.1	11.8	11.8

β, response criterion set by the participant: lower values represent more liberal criteria; d', sensory sensitivity: higher values represent higher sensitivity, infinite values mean perfect discriminability; FA, false-alarm rate (% of responses to catch-trials)

auditory targets (311 ± 33 ms) was even slightly higher than mean RT of invalidly cues auditory targets (307 ± 29 ms). The cueing effects in the audiovisual signals condition were significant only for visual-first conditions and the synchronous target condition A(0)V (after FDR-correction for multiple comparisons; see also Table 5). No significant effect of target modality (MRT_A–MRT_V) was obtained ($t(7) = -2.006$, $p = 0.085$ and $t(7) = 1.729$, $p = 0.127$ for invalid and valid cue condition, respectively). In general, the mean RT curve of SOA-dependent targets followed the typical wing-shaped pattern in both validity conditions—that is, responses to synchronous redundant targets were the fastest, irrespective of cue condition (Table 5). With increasing SOA, mean RT increased and approached the unimodal target mean RT at both ends. Unlike Experiment 1, the cueing effects were not weakest in the redundant synchronous condition with cueing effects increasing with SOA. Here, the cueing effect varied systematically with the onset of the visual target: The cueing effect was most pronounced in the visual unimodal condition and became smaller, the more the visual target component was delayed (with respect to the auditory target component) and completely disappeared in the auditory unimodal condition (see t values in Table 5).

Race-model inequality

The observed redundancy gains were again significantly greater than a race model would predict in both invalid cue conditions ($T_{max} = 3.00$; critical $T^*_{max} = 2.72$; $p = 0.014$) and valid cue conditions ($T_{max} = 2.31$; critical $T^*_{max} = 2.25$; $p = 0.039$). Thus, unlike Experiment 1, significant violations of race-model predictions were obtained for both valid and invalid cue conditions.

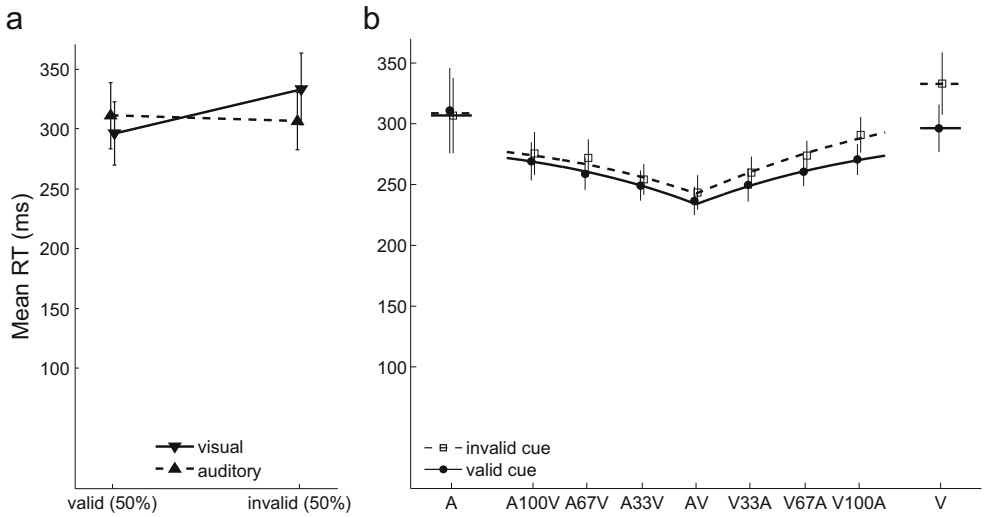

Fig. 3 Experiment 2—**a**: Mean response time (± 95% confidence intervals) of auditory and visual targets following valid and invalid cues. The visual cueing effect and the modality effect in the invalid cue condition were significantly non-zero. **b**: Mean response times observed in valid and invalid cue trials averaged across all participants ($n = 8$) together with averaged DSM predictions of the model with modality-specific cueing effects. Error bars denote averaged estimates of the 95% confidence intervals. As shown, the model could explain the observed patterns of results in Experiment 2. Note that there was no cueing effect in the unimodal auditory condition (A) and that the cueing effect again systematically varies with the SOA: the more the visual target is delayed (left half of the curve) the smaller was the observed effect of cue validity

Table 5 Experiment 2—Mean (*MRT*, ms) and standard deviations (*SD*) of response times to auditory (A), visual (V), and audiovisual (AV) targets presented with onset asynchrony after valid cues and invalid cues

Target condition	Cue condition				Cue validity effect: $MRT_{invalid} - MRT_{valid}$	
	Valid		Invalid			
	MRT	*SD*	*MRT*	*SD*	*t*(7)	*p*
A	311	33	307	29	–0.568	0.588
A100V	269	19	276	23	1.145	0.290
A67V	259	20	272	23	3.490	0.010*
A33V	249	22	254	20	2.478	0.042
AV	237	23	244	23	4.842	0.002*
V33A	250	24	260	28	3.549	0.009*
V67A	261	20	274	22	6.244	<0.001*
V100A	271	19	291	24	6.225	<0.001*
V	296	32	333	36	6.715	<0.001*

Results are based on data from 8 participants. Participant specific mean RTs were retrieved from RT distributions corrected for anticipations (see *Methods*). For valid cues, these distributions included 96 responses per participant; for invalid cues also 96 responses (50% cue validity). The reported *p* values result from two-sided, paired *t* tests for a cueing effect ($MRT_{invalid} - MRT_{valid}$) in the respective SOA condition. Significant differences (after FDR-correction) are marked with an asterisk. The *MRT* data corresponds to the data points depicted in Fig. 3a, b.

Diffusion superposition model

The aggregate model fit on group level revealed that the modality-specific attention model fitted best to data ($X^2 = 83.36$; $df = 88$; $p = 0.620$; Fig. 3b). In clear contrast to Experiment 1, the restriction of a modality-independent attentional effect leads to considerably increased X^2, that is, worse model fits ($X^2 = 122.0$; $df = 96$; $p = 0.038$). Both models correctly described the asymmetric wing-shaped pattern of mean RT in both the valid and the invalid cue condition, but it became evident that the modality-invariant model systematically underestimated cueing effects in visual targets and overestimated cueing effects in auditory targets. The modality-specific attention model well captured the difference in the cueing effect between auditory and visual unimodal target trials (Fig. 3b). The comparison between these two models thus yielded a significant difference ($\Delta X^2 = 38.67$; $df = 8$; $p < 0.001$), that is, the attention factors were significantly different across modalities (Table 6): the effect of valid cues on auditory processing ($g_A = 1.02$) was much lower than on visual processing ($g_V = 1.28$). Post-hoc we tested the auditory attention parameter g_A against unity and obtained a nonsignificant difference ($\Delta X^2 = 4.86$; $df = 8$; $p = 0.772$). This implies that there was, if at all, a negligible cueing effect on processing in the auditory channel. Unlike Experiment 1, the estimated drift for processing in the auditory channel was similar to that for visual processing (Table 6). The effect of cueing on the mean latency of nonperceptual processes (μ_M) was again negligible ($\Delta X^2 = 10.46$; $df = 8$; $p = 0.234$). In sum, the observed data in Experiment 2 are best described by the modality-specific cueing effects model, because visual exogenous cues effectively facilitated visual processing but had minimal effects on auditory processing. The modality-invariant model could not explain the observed pattern of response times and was rejected based on both qualitative and quantitative assessment.

Discussion

In Experiment 2, we investigated the effects of exogenous cues on the performance in a multisensory redundant signals task. The effects were similar to endogenous cueing effects

Table 6 Experiment 2—Parameter estimates and model fit of the Diffusion Superposition Model with modality-specific cueing effects

Parameter	Average	SD
μ_A	0.72	0.11
σ_A	13.3	9.2
μ_V	0.63	0.12
σ_V	6.9	3.3
g_A	1.02	0.06
g_V	1.28	0.11
μ_M	167.2	22.2
Goodness of fit (*higher results indicate worse fit*)		
$X^2(88)$	83.36	$p = 0.620$
Comparison with modality invariant cueing effects		
$\Delta X^2(8)$	38.67	$p < 0.001$

μ_A, σ_A, μ_V, σ_V: drift and diffusion constant for auditory and visual diffusion process; g_A, g_V: scale factor (facilitation) for validly cued targets in the two modalities; μ_M: mean latency of residual processes

but also differed in several important aspects. As expected, the cues produced significant cueing effects in the visual unimodal targets (Jonides, 1981; Müller & Rabbitt, 1989) and nonsignificant cueing effects for auditory targets (Buchtel & Butter, 1988; Spence & Driver, 1997). Here, we extended the cross-modal cueing setup to a redundant signals experiment and used a computational model to assess redundancy gains and cue validity effects in a single model. The race model was rejected because redundancy gains in both the valid and the invalid cue condition were greater than predicted by parallel-first terminating processing. We fitted a coactivation model (Schwarz, 1994) to the data and tested for different assumptions regarding the mechanism of attentional modulation by exogenous cues. In Experiment 1, the modality-invariant cueing model best described both redundancy gains and cueing effects. In contrast, in Experiment 2, the modality-specific cueing model accounted much better for the observed patterns of results. The smallest cue validity effects were found in the auditory single target condition.

Considering the parameter estimates and data, the attention effects on auditory processing were small or even absent. This was evident in the comparison of validly and invalidly cued auditory targets that showed no cueing effect. Moreover, if one compares the model parameters estimated from the data of Experiment 2 with those obtained in Experiment 1, it is striking to see how inefficient auditory processing became with the same targets. The only difference is that exogenous instead of endogenous cues were used in Experiment 2. The perception of both the auditory stimuli (unimodal targets) *and* the auditory stimulus components (bimodal targets) hardly benefited from prior cueing. To some extent, this is quite the opposite of what we observed in Experiment 1. Whereas endogenous cues lead to facilitation effects of the whole multisensory percept, exogenous cues facilitate only visual targets or the visual target component. The perception of an accompanying auditory target component is not facilitated. Therefore, in stimulus-driven attention visual cues seem to facilitate mainly, or, even exclusively, the visual perception of audiovisual signals and their integration into a common percept. Alternatively, the spatial properties of visual exogenous cues do not seem to play a significant role in auditory perception. While auditory signals can facilitate visual processing even if they are location-unspecific (Van der Burg et al., 2008), possibly due to effects of stimulus driven multisensory integration on spatial attention (Talsma et al., 2010), this seemingly does not generally apply to visual cues in an auditory target detection task. Note, however, that other studies have obtained evidence for supramodal attention also with stimulus-driven attention (McDonald, Teder-Sälejärvi, Heraldez & Hillyard, 2001). One difference between their study and ours that might be critical to explain the different results is the spatial separation between the visual cue and the auditory target (component). When spatial separation is close to zero (e.g., LEDs attached to loudspeakers), responses to an auditory target are influenced by cue validity. This interpretation has been discussed before (Spence et al., 2004) and would explain the reported null-effects of visual cues on auditory targets (Buchtel & Butter, 1988; Spence & Driver, 1997) when there is some degree of spatial separation between visual cue and auditory target. However, in addition to these studies that reported null-effects of visual exogenous cues on auditory target detection (Buchtel & Butter, 1988) or auditory target elevation discrimination (Spence & Driver, 1997), the results of Experiment 2 suggest that this also is the case for auditory target components in redundant bimodal targets. This point is critical, because the interpretation of similar results obtained so far often hinges on the nonsignificant test result of a single experimental condition. The model parameters on which we base our interpretation incorporate data of not less than eight conditions (all invalid conditions except the visual unimodal target condition that is independent of the auditory attention factor). It is worth noting that in all models, the mean latency of residual processes was kept constant, irrespective of modality, redundancy, and cue validity. Thus, we conclude that exogenous cues influence (visual) perception, at least much more than nonperceptual processes, in the integration of audiovisual signals.

Experiment 3

Experiments 1 and 2 differed in three aspects of the employed cues: the type of the cue (central/symbolic vs. peripheral/nonsymbolic), the validity of the cue (predictive vs. nonpredictive) and the cueing interval. In Experiment 3, we tested whether the difference in cueing effects observed in Experiments 1 and 2 could be explained by the cue validity. Experiment 3 was almost identical to that of Experiment 2, with the critical difference that the peripheral cues had the same cue validity (75%) as in Experiment 1.

Methods

Participants

Fourteen new participants (mean age: 25.7 [range 21–30] years; 9 females, and 3 left-handed) were tested in Experiment 3. Before participation, they gave written, informed consent. All were naïve regarding the purpose of the experiment. All participants reported normal or corrected-to-normal visual acuity and normal hearing.

Table 7 Experiment 3—Results of signal detection theory parameter estimation and frequency of false alarms

	Participant													
	1	2	3	4	5	6	7	8	9	10	11	**12**	13	14
β	2.20	1.97	1.43	1.69	1.56	1.97	1.43	2.31	1.43	1.91	1.59	**0.44**	1.86	1.77
d'_{valid}	4.15	5.14	4.26	4.86	3.31	5.14	3.68	5.48	4.39	4.87	∞	∞	4.28	4.51
$d'_{invalid}$	4.66	∞	4.26	∞	3.25	∞	3.79	5.14	4.03	∞	4.43	**3.27**	4.32	4.37
FA	1.4	2.4	7.6	4.5	5.9	7.6	7.6	1.0	7.6	2.8	5.6	**33.0**	3.1	3.8

β, criterion set by the participant: lower values represent more liberal criteria; d', sensory sensitivity: higher values represent better sensitivity, infinite values mean perfect discriminability; FA, false-alarm rate (% of responses to catch-trials)

Task and stimuli

In Experiment 3, we used nonsymbolic peripheral cues encompassing one of the two possible target locations. In contrast to Experiment 2, these cues were predictive in that the subsequent target (if any, see Experiment 1 for a description of catch trials) appeared in three of four cases at the position framed by the cue (75% cue validity). All other parameters of the setup and the experimental task were identical to Experiment 2.

Procedure

As in the previous experiments, each participant was tested in three separate sessions and each session lasted approximately 45 minutes. The number of trials is again the same as in the previous experiments; due to the cue validity we obtained 144 replications for each valid and 48 for each invalid cue conditions (Experiment 1).

Statistical analyses

The analyses were the same as in Experiments 1 and 2. Again, due to changed cue validity, the kill-the-twin procedure was employed with weights 4.5 and 1.5 for valid cue trials and invalid cue trials, respectively. Slow responses were again rare and accounted for < 1% of RT in most participants (1.2%, 3.1, and 4.2% in Participants 7, 1, and 5, respectively). Otherwise, the same tests and model fits were conducted as described in the previous sections.

Results

Analysis of response times and anticipations

Responses to catch trials were more frequent than in Experiment 2, but less frequent than in Experiment 1. Except for one participant, response frequency to catch trials was between 1% and 10%. This also is evident in the estimated response criteria of signal detection theory (Table 7). One participant responded to 30% of the catch trials (β = 0.44) and, by the same reasoning as before, was excluded from further analysis. The ANOVA results, race-model tests, and the assessment of model fits therefore are based on 13 participants.

The ANOVA yielded significant main effects for the two factors Cue Validity ($F(1,12) = 32.74$, $p < 0.001$) and SOA ($F(1,11) = 28.27$, $p < 0.001$) but not for the interaction Cue Validity × SOA ($F(1,11) = 3.359$, $p = 0.092$). In line with Experiments 1 and 2, we found a cueing effect of visual exogenous cues on the mean response times of visual targets (Table 8; Fig. 4a): mean response time (± *SD*) in the visual target condition decreased from 354 ms(± 89 ms) to 312 ms (± 63 ms), if targets were presented at the cued location rather than the opposite location ($t(12) = –4.741$, $p < 0.001$). Like Experiment 2, there was no significant cueing effect for auditory targets ($t(12) = –0.730$, $p = 0.479$): the mean RT of validly cued auditory targets (310 ± 87 ms) was only slightly lower than the mean RT of invalidly cues auditory targets (318 ± 78 ms). Also in line with Experiment 2, we only found a modality effect (MRT_A–MRT_V) in the invalid cue condition ($t(12) = –3.709$, $p = 0.003$) but not in the valid cue condition ($t(12) = –0.085$, $p = 0.934$). The mean RT curve of SOA-dependent targets followed the expected wing-shaped pattern in both validity conditions and the cueing effect on the audiovisual targets again varied systematically with SOA. The cueing effect became smaller with increasing delay of the visual target compared with the preceding auditory target, but the decrease was less pronounced than in Experiment 2: only in the auditory unimodal condition it is evident that the cueing effect is absent or, negligible small (Table 8).

Race-model inequality

The race model inequality was again violated in both the valid condition ($T_{max} = 3.545$; $T^*_{max} = 2.305$; $p = 0.003$) and the invalid cue condition ($T_{max} = 3.351$; $T^*_{max} = 2.257$; $p = 0.005$), ruling out separate activation models from consideration.

Table 8 Experiment 3—Mean (*MRT*, ms) and standard deviations (*SD*) of response times for auditory (A), visual (V), and audiovisual (AV) targets presented with onset asynchrony after valid cues and invalid cues

Target condition	Cue condition				Cue validity effect: $MRT_{invalid} - MRT_{valid}$	
	Valid		Invalid			
	MRT	*SD*	*MRT*	*SD*	*t*(12)	*p*
A	311	87	318	78	0.730	0.479
A100V	267	52	281	61	3.252	0.007*
A67V	260	52	278	56	3.528	0.004*
A33V	251	47	265	52	2.659	0.021*
AV	237	42	259	52	5.418	<0.001*
V33A	259	42	279	50	4.246	0.001*
V67A	271	42	294	54	3.568	0.004*
V100A	285	42	305	49	4.990	<0.001*
V	312	63	354	89	4.741	<0.001*

Results are based on data from $n = 13$ participants. Participant specific mean RTs were retrieved from RT distributions corrected for anticipations (see *Methods*). For valid cues, these distributions included 144 responses per participant; for invalid cues 48 responses (75% cue validity). The last two columns are results from two-sided paired *t* tests for a cueing effect ($MRT_{invalid} - MRT_{valid}$) in the respective SOA condition. Significant differences (after FDR-correction) are marked with an asterisk. The *MRT* values correspond to the data points depicted in Fig. 4a, b

Diffusion superposition model

The same models as described before were fitted to the reaction times of all participants. In line with the ANOVA results, the null-model assuming no cueing effects fitted poorly to the observed data ($X^2 = 642.3$; $df = 169$; $p < 0.001$). Also, the fit of the modality-invariant model was poor ($X^2 = 233.0$; $df = 156$; $p < 0.001$). Only the model for modality-specific attention effects provided an adequate fit to the observed data ($X^2 = 179.9$; $df = 143$; $p = 0.020$). One participant showed clear evidence for nonperceptual cueing effects ($\Delta X^2 = 11.76$; $df = 1$; $p < 0.001$); if μ_M was allowed to vary across invalid and valid cue condition for that participant, the overall fit of the modality-specific cueing effects model improved substantially ($X^2 = 168.2$; $df = 142$; $p = 0.066$). As in Experiment 2, the difference between the modality-invariant attention model and the modality-specific attention model was

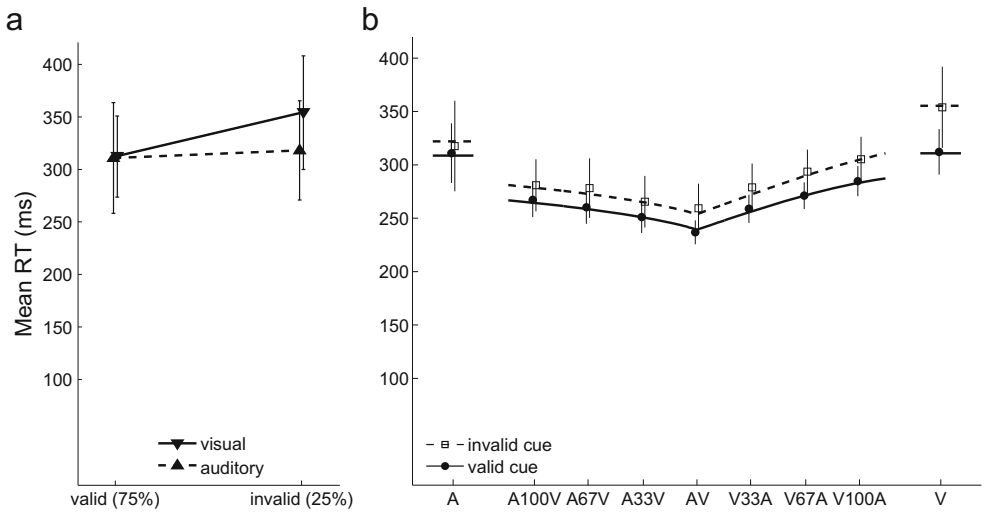

Fig. 4 Experiment 3—**a**: Mean response time (± 95% confidence intervals) of auditory and visual targets following valid and invalid cues. The visual-cueing effect and the modality effect in the invalid cue condition were significantly non-zero. **b**: Mean response times observed in valid and invalid cue trials averaged across all participants ($n = 13$) together with averaged DSM predictions of the model with modality-specific cueing effects. Error bars denote averaged estimates of the 95% confidence intervals. As shown, the model could explain the observed patterns of results in Experiment 3, assuming modality-dependent attention effects. Again, there was no cueing effect in the auditory unimodal condition (A)

significant (ΔX^2 = 53.10; df = 13; $p < 0.001$); the data again could best be described by the modality-specific attention model. Contrary to Experiment 2, the auditory attention factor (g_A = 1.16) was significantly different from unity (ΔX^2 = 64.14; df = 13; $p < 0.001$) but still smaller than the visual attention factor (g_V = 1.34). As such, the result of Experiment 3 is located between the two rather extreme results of model fit results obtained in Experiments 1 and 2 (Tables 3, 6, and 9). Possible explanations for this intermediate kind of cross-modal attentional facilitation are discussed in the next section.

Discussion

The purpose of Experiment 3 was to test whether the observed differences in cueing effects between endogenous and exogenous cues on multisensory integration were due to the difference in cue validity. Particularly, we wanted to test whether the absence of a cueing effect on auditory processing in Experiment 2 is explained by the noncontingent cue-target relation of the exogenous cues (Kristjánsson & Nakayama, 2003). Therefore, we repeated Experiment 2 with predictive cues to investigate the effect of a cue-target contingency on the performance in the audiovisual redundant signals task. As explained above, exogenous attention was investigated by nonpredictive cues to preclude the possibility that voluntary (endogenous) controlled attention contributes to the direction of spatial attention and the perception of signals (Spence & Driver, 1996). This distinction seemed to work well, as we observed quite different effects of cross-modal cueing effects in the Experiments 1 and 2.

However, due to the various differences in the cueing parameters of Experiments 1 and 2, the two spatial cueing experiments are somewhat hard to compare, because it remains unclear if the differences in cross-modal cueing effects are due to the change in cue validity, the different forms of the cues (i.e., symbolic vs. non-symbolic), or the different cueing interval. Experiment 3 served as to control for the change in cue validity between Experiment 1 and 2. In line with the conclusions of Experiment 2, the exogenous cues (peripheral, nonsymbolic) exerted modality-specific attention effects: Processing in the visual channel was again more facilitated by the cues than processing in the auditory channel. However, the results differ from Experiment 2 insofar that there was at least some extent of facilitation of auditory processing (i.e., across senses). We believe that in Experiment 3 both exogenous and endogenous attention were at work. The peripheral cues attracted attention in a bottom-up manner, but the predictive value of the cue motivated participants to direct spatial attention to that location in a top-down (cross-modal) fashion. The endogenous effects were rather weak, supposedly because the short cueing interval (100 ms) does not allow for efficient endogenous deployment of spatial attention. We would expect similar differences regarding the different cueing interval between Experiments 1 and 2. An interplay between endogenous and exogenous attention has already been discussed by Spence and Driver (1996), who suggested using nonpredictive cues to minimize effects of endogenous attention (for a more recent view on the dichotomy of endogenous and exogenous attention, see Awh, Belopolsky & Theeuwes, 2012). We conclude that the change in cue validity was not the decisive factor to explain the different cross-modal cueing effects, as we found strong evidence for modality-specific facilitation effect also for *predictive* exogenous cues. However, endogenous attention seemed to be in effect due to the cues being predictive, so that the (modality-invariant) endogenous attention effects also facilitated processing in the auditory channel, albeit to a much lesser extent than the endogenous cues in Experiment 1.

General discussion

While many studies have reported cross-modal cueing effects of visual cues on auditory perception in endogenous cueing experiments (Farah et al. 1989; Spence & Driver, 1996; Eimer & Schröger 1998), the results of studies involving exogenous cues are rather inconclusive (Buchtel & Butter, 1988; Spence & Driver, 1997; Ward et al. 2000). In three experiments, we employed both types of cues in an audiovisual redundant signals task. This extension investigated the effects of spatial attention on the integration of audiovisual signals (Bertelson et al., 2000; Vroomen et al., 2001). With redundant audiovisual signals one cannot only investigate cross-modal cueing effects (at fine-grained resolution due to SOA-variation) but also how the processing of auditory and visual components of a redundant signal takes place and how this processing is modulated by attention. We addressed this question with the application of a diffusion model to test explicitly for differential cross-modal cueing effects. The model has repeatedly shown to predict redundancy gains successfully in bimodal divided attention tasks (Schwarz, 1994; Diederich, 1995; Gondan, Götze & Greenlee, 2010; Gondan et al., 2011; Blurton, Greenlee & Gondan, 2014). On the one hand, the inspection of the model parameters supports previous interpretations and extends results in endogenous cueing. On the other hand, it allows for interesting new interpretations of exogenous cross-modal cueing.

As expected, we observed strong cross-modal cueing effects under endogenous control of attention (Experiment 1). The interesting point is that cross-modal cueing effects were not only observed with auditory targets. Rather, all audiovisual targets exhibited effects of valid and invalid cueing (Table 2). Under all conditions, cueing effects were in agreement with a modality-invariant supramodal attentional system: Cueing effects were not only cross-modal in nature, but

Table 9 Experiment 3—Parameter estimates and model fit of the Diffusion Superposition Model with modality-specific cueing effects

Parameter	Average	SD
μ_A	0.79	0.30
σ_A	20.7	16.5
μ_V	0.62	0.21
σ_V	4.0	2.5
g_A	1.16	0.21
g_V	1.34	0.24
μ_M	172.6	30.3
Goodness of fit (*higher results indicate worse fit*)		
$X^2(143)$	179.9	$p = 0.020$
Comparison with modality invariant cueing effects		
$\Delta X^2(13)$	53.10	$p < 0.001$

μ_A, σ_A, μ_V, σ_V: drift rate and diffusion constant for auditory and visual diffusion process; g_A, g_V: scale factor (facilitation) for validly cued targets in the two modalities; μ_M: mean latency of residual processes

even equal in strength across modalities. This notion of attentional cueing effects is compatible with modality-specific differences in mean RTs for unimodal auditory and visual targets. While differences in mean RTs for auditory and visual stimuli suggest modality-specific cueing effects, the application of the diffusion model and the interpretation of its parameters suggest otherwise. In the diffusion model, perceptual facilitation operates through attention parameters that describe processing in both channels for stimuli that fall into the focus of attention. According to the model, differences in cueing effects between audition and vision follow from the fact that auditory signals are processed faster than visual signals.

Top-down control of spatial attention seems, therefore, to be largely compatible with a common, supramodal attention system (Talsma et al., 2010; Koelewijn et al., 2010). The present finding of a modality-invariant cueing effect is in line with previous studies on voluntarily controlled attention and its effect on multisensory processing (Farah et al., 1989; Talsma et al., 2010). The results also are consistent with effects of sustained spatial attention in multisensory integration of audiovisual signals (Gondan et al., 2011).

Based on the data of the present study, it cannot be decided whether voluntarily controlled attention facilitates the perception of already integrated multisensory objects (preattentive/early integration) or if attention facilitates the integration of unimodal percepts into a single object (late integration). Neurophysiological studies support both notions: The effects of attention on early processes can be the consequence of preattentive integration (Eimer, 2001). However, it also has been demonstrated that attention modulates event-related scalp potentials of redundant signals at various stages, including late processes (Talsma & Woldorff, 2005).

It has been argued (Spence & Driver, 1997) that cross-modal attention of visual cues on auditory targets depends on late processes that are not elicited in simple response tasks, such as those used in the present series of experiments. This argument addresses the representation of auditory and visual information in the brain. Vision is spatiotopic from the onset of signal processing, where different neuronal populations encode information at different spatial locations. Audition, on the other hand, is not spatiotopically but rather tonotopically organized. Spatial location has to be extracted from different cues, for example, interaural-time or level differences, the latter related to sound attenuation caused by the skull, whereas the former by different distance between the sound source and the left or right ears. This extraction takes place at a rather late stage of auditory processing and that spatial information therefore might not be available in a simple response task. The difference between auditory and visual modality seems not to be crucial for endogenous cross-modal cueing effects, although as endogenous visual-on-auditory cueing effect have been reported in a perceptual decision task (Spence & Driver, 1996), a detection task (Farah et al., 1989), and now in a redundant signals task. The results of the present study demonstrate that the facilitation effects of attention on the processing of objects with features from different senses are largely independent of modality. In line with this, we interpret the observed attention effects in Experiment 1 as evidence for facilitation of already integrated multisensory percepts (early integration).

Experiments 2 and 3 used exogenous control of attention. At first glance, the findings add to the ambiguity created by results of studies on exogenous cross-modal (i.e., visual-to-auditory) cueing. Some studies have found cueing effects, whereas others did not and methodological differences between those studies hardly allow for clear-cut conclusions (for review, Spence et al., 2004). The consensus is that visual cues influence auditory perception, but only under certain conditions that are not yet fully specified. A number of arguments have been put forward that aim at explaining the conditions under which cross-modal cueing occurs. The arguments include cue validity (predictive vs. nonpredictive), response task, differences in spatial resolution between modalities, and criterion shifts (Spence et al., 2004; Koelewijn et al., 2010). For example, it has been argued that exogenous visual cues lead to cueing effects, if they are predictive. This is supported by the results of Experiment 3 in which we found weak cross-modal cueing effects in the model parameters (i.e., $g \neq 0$) but not mean RT (i.e., $MRT_{valid} \approx MRT_{invalid}$ in Condition A).

It also has been put forward that criterion shifts rather than facilitation by attention account for the faster responses to targets presented at the precued location. The diffusion model

is mute with respect to this distinction. The attention parameter can reflect improved processing (g) or a criterion shift ($1/g$). This concern applies not specifically to cross-modal attention studies, but to all studies that are based on the original spatial cueing task (Posner et al., 1978; Posner, 1980; Jonides, 1981). A perceptual decision task can help to distinguish criterion shifts from attention effects, for example, by application of the signal detection model (Green & Swets, 1966). We used a detection task and applied the response time model to the resultant data. The signal detection model has been connected recently to diffusion models that not only dissociate criterion and sensitivity but also take response times into account (Wagenmakers, van der Maas & Grasman, 2007). In future studies, a combination of the DSM and diffusion models for discrimination tasks could help to further elucidate cross-modal relationships.

One great advantage of response-time models, such as the DSM, is that they allow for detailed analyses of response times across a large number of experimental conditions. In addition to the differential effects of exogenous and endogenous cueing on response time, we also obtained evidence that indicate that perceptual processes were facilitated or attenuated, depending on the validity of the cue. If nonperceptual processes (μ_M) accounted for the results, one would expect that also the latency of these processes would be subject to change under different cue validities. This was not observed in the present study; rather, the redundancy gains observed across different type of cues indicate that it was perceptual processing that was facilitated when attention was allocated to the target location. Critically, with exogenous cues only the processing of visual features was facilitated—regardless of whether the cue was predictive or not. The detection process of auditory features was largely constant, irrespective of the preceding cue. In a recent review, Talsma et al. (2010) integrated the diverse literature of cross-modal cueing effects on multisensory integration into a common framework. They reviewed studies involving top-down as well as bottom-up control of cross-modal attention and propose a model for both forms of attentional control. According to this model, auditory signals presented together with a visual target lead to bottom-up induced audiovisual integration, which direct visuospatial attention. This conclusion was based on the finding that spatially uninformative auditory stimuli can produce location unspecific facilitation effects on visual processing in a visual search task (Van der Burg et al., 2008).

On this note, the data observed in Experiment 2 could be the result of visual cues leading to cross-modal facilitation of (auditory) perception, irrespective of the location of the subsequent (auditory) target because of the low spatial resolution of audition (Spence & Driver, 1994). This would not only explain the null effects of visual cues on the latency of auditory target perception but also the null effect on auditory processing in all audiovisual targets, including the redundant signals conditions. The comparison of response times and the estimated model parameter from those data suggest an alternative interpretation: both mean response times as well as parameter estimates support the notion that auditory targets were never attended to, regardless of whether the preceding visual cue was indicating the correct location or the opposite location. The visual modality would then dominate processing of redundant signals; this dominance would have been induced by the visual cue that precedes the auditory one. If visual processing indeed had dominated the perception of audiovisual signals after exogenous (but not endogenous) cues, the greater spatial resolution of the visual system might have differentially facilitated the processing of spatially separated visual and auditory targets (Spence et al., 2004; Koelewijn et al., 2010). This would explain why the processing of the same auditory targets presented at the same locations was effectively facilitated after valid endogenous cues, but not after valid exogenous cues. Of course, this interpretation needs to be confirmed by data, because, to our knowledge, the criticality of spatial separation of auditory and visual targets in exogenous (but not by endogenous attention) often has been discussed but never directly investigated (Prime, McDonald, Green & Ward, 2008, for spatial influences on exogenous cross-modal attention).

As such, the results from the redundant audiovisual conditions support existing theories on endogenous cross-modal interactions and shed new light on exogenous cross-modal interactions. Future studies on exogenous links between vision and audition can help to refine our knowledge further about the relationship between these two sensory systems.

Acknowledgement We would like to thank Francesco Marini, Durk Talsma, and an anonymous reviewer for helpful comments on an earlier version of this manuscript. Supported by grant GR 988/20-2 from the German Research Foundation (Deutsche Forschungsgemeinschaft) and by the Dynamical Systems Interdisciplinary Network, University of Copenhagen.

Appendix

In this appendix, the DSM and the model comparisons presented in the main text are described in more detail. In the DSM, it is assumed that the information accumulation can be described formally by a Wiener process with drift μ and diffusion constant σ in the presence of an absorbing barrier c (i.e., the response criterion). The process is time-homogenous and its first passage distribution is well-known (inverse Gaussian). Particularly, the mean first passage time is $E(\mathbf{D}) = c/\mu$. In the model, it is assumed that both auditory and visual modalities are processed in separate channels, described by separate drift and variance. In the case of redundant signals, it is assumed that the contributions of the modality-specific channels are added into a common channel: $\mathbf{X}_{AV}(t) = \mathbf{X}_A(t) + \mathbf{X}_V(t)$. This can best be seen in the case of

synchronous presented redundant signals, for which the mean detection time is then $E(\mathbf{D}) = c/(\mu_A+\mu_V)$. That is, the redundancy effect is explained by the assumption that in the case of redundant signals (AV), more information is pooled into the common channel than in the case of any unimodal stimulus. It also is easily seen that the expected value for **D** decreases, if either the criterion c is decreased or, equivalently, the modality-specific drifts are multiplied by a constant $g < 1$. This decrease affects the mean detection times of all conditions, including those of redundant signals presented with SOA (see Schwarz, 1994, Eq.10 for details). As stated in the main text, this part of the model describes stimulus detection times, not response times. It is further assumed that response time consists of this detection process **D** and nonperceptual processes **M** ("motor processes"). The mean latency of all nonperceptual processes is represented in a separate, single parameter (μ_M).

The first model tested is the DSM as described by Schwarz (1994). Both cue validity conditions are described by the same set of parameters. This constitutes the null-model, which served for testing cue validity effects across all SOA-conditions. The model had five free parameters (μ_A, σ_A, μ_V, σ_V, μ_M) that need to be estimated from data to fit the model to a set of observed mean response times.

The second model contains an additional parameter for the valid cue conditions. In this model, attention effects are modelled with a scaling parameter g: $\mu_{valid} = \mu \times g$ and $\sigma_{valid} = \sigma \times g$. The effect of this additional parameter is again best illustrated by the respective mean detection time: $E(\mathbf{D}|\text{valid cue}) = c / (\mu \times g)$. If g is greater than unity, the expected detection time $E(\mathbf{D})$ (and predicted mean reaction time) decreases. If g is unity, the predictions for targets at expected locations are identical to those presented at unexpected locations (i.e., the null-model). This holds for all stimulus conditions following a valid cue, including redundant targets presented with SOA. As the cueing effect is allowed to affect only this single parameter, the underlying assumption is a modality-invariant attention effect. This is a strong restraint and, given the diverging evidence in cross-modal cueing tasks, may or may not hold in a given task and data set. This model had six free parameters (μ_A, σ_A, μ_V, σ_V, g, μ_M) that need to be estimated for each participant.

To test the modality-invariance assumption, we fitted a third model to the data of all experiments, which allowed for modality-specific attention effects to be included in the model. The derivation is straightforward: instead of using a single parameter for attention effects (g), we added separate scaling parameters for each modality (g_A, g_V). This model is connected to the modality-invariant model by the constraint $g_A = g_V$. Without this constraint, the model is able to predict mean response times under the assumption of modality-specific cueing effects. If, for example, auditory targets are more efficiently processed at expected locations, g_A should be greater than unity and the model predicts an decreased processing time in the auditory channel: $E(\mathbf{D}_A|\text{valid cue}) = c/(\mu_A \times g_A)$. Critically, and in contrast to the modality-invariant attention model with one attention parameter, this model allows for differential effects of spatial cueing in the processing channels of both modalities. This model had seven free parameters (μ_A, σ_A, g_A, μ_V, σ_V, g_V, μ_M).

Because all models are nested, we were able to test for the necessity of the added parameters by means of X^2-difference tests. For these tests, we calculated the difference in model fit between two models. This statistic is approximately χ^2-distributed with *df* determined by the difference of the number of free parameters. In the model comparisons on group level as presented in the main text (Fig. 1c), *df* corresponds to the number of participants, because in each model comparison we reduced the number of parameters by one for each participant: the degrees of freedom for each participant's model fit was $df = 18 - 5 = 13$ for the null-model (predicting no cueing effects), $df = 12$ for the model with modality-invariant cueing effects and $df = 11$ for the model with modality-specific cueing effects. Parameters were deemed as significantly different, if the difference in model fit was greater than the 95%-quantile of the respective χ^2-distribution.

In all three models, the residual latency parameter μ_M was restricted to be constant across all conditions. If the latency of nonperceptual processes does not depend on attentional manipulations, this can be taken as evidence for a perceptual interpretation of cueing effects. Formally, this adds another constraint to the model; the appropriateness of this constraint was post-hoc tested in a likewise nested model comparison.

References

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in Cognitive Sciences, 16,* 437–443.

Bashinski, H. S., & Bacharach, V. R. (1980). Enhancement of perceptual selectivity as the result of selectively attending to spatial locations. *Perception & Psychophysics*, *28*, 241–248.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B, 57,* 289–300.

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, *62*, 321–332.

Blurton, S. P., Greenlee, M. W., & Gondan, M. (2014). Multisensory processing of redundant information in go/no-go and choice responses. *Attention, Perception, & Psychophysics, 76,* 1212–1233.

Buchtel, H. A., & Butter, C. M. (1988). Spatial attention shifts: implications for the role of polysensory mechanisms. *Neuropsychologia*, *26*, 499–509.

Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: support for signal enhancement. *Vision Research, 40,* 1203–1215.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience, 7,* 308–313.

Dosher, B. A., & Lu, Z.-L. (2000). Noise exclusion in spatial attention. *Psychological Science, 11,* 139–146.

Diederich, A. (1995). Intersensory facilitation of reaction time: evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology, 39,* 197–215.

Diederich, A., & Colonius, H. (1987). Intersensory facilitation in the motor component? *Psychological Research, 49,* 23–29.

Driver, J., & Spence, C. (1998). Attention and the cross-modal construction of space. *Trends in Cognitive Sciences, 2,* 254–262.

Eimer, M. (2001). Crossmodal links in spatial attention between vision, audition and touch: evidence from event-related brain potentials. *Neuropsychologia, 39,* 1292–1303.

Eimer, M., & Schröger, E. (1998). ERP effects of intermodal attention and cross-modal links in spatial attention. *Psychophysiology*, *35*, 313–327.

Eriksen, C. W. (1988). A source of error in attempts to distinguish coactivation from separate activation in the perception of redundant targets. *Perception & Psychophysics, 44,* 191–193.

Fairhall, S. L., & Macaluso, E. (2009). Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *European Journal of Neuroscience, 29,* 1247–1257.

Farah, M. J., Wong, A. B., Monheit, M. A., & Morrow, L. M. (1989). Parietal lobe mechanisms of spatial attention: modality specific or supramodal? *Neuropsychologia, 27,* 461–470.

Feintuch, U., & Cohen, A. (2002). Visual attention and coactivation of response decisions for freatures from different dimensions. *Psychological Science, 13,* 361–369.

Gondan, M. (2009). Testing the race model inequality in redundant stimuli with variable onset asychrony. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 575–579.

Gondan, M. (2010). A permutation test for the race model inequality. *Behavior Research Methods*, *42*, 23–28.

Gondan, M., Blurton, S. P., Hughes, F., & Greenlee, M. W. (2011). Effects of spatial and selective attention on basic multisensory integration. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 1887–1897.

Gondan, M., & Fimm, B. (2013). Simple correction methods for task completion times contaminated by errors, outliers and omissions. Paper presented at the annual meeting of the Society of Mathematical Psychology, Potsdam, Germany.

Gondan, M., Götze, C., & Greenlee, M. W. (2010). Redundancy gains in simple responses and go/no-go tasks. *Attention, Perception, & Psychophysics, 72,* 1692–1709.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology, 63,* 289–293.

Jonides, J. (1981). Voluntary vs. automatic control over the mind's eye's movement. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 187–203). Hillsdale, NJ: Erlbaum Associates Inc.

Kalbfleisch, J. D., & Prentice, R. L. (2002). *The statistical analysis of failure time data*. New York: Wiley.

Kiesel, A., Miller, J., & Ulrich, R. (2007). Systematic biases and Type I error accumulation in tests of the race model inequality. *Behavior Research Methods, 39,* 539–551.

Koch, S. H., Weir, C., Westenskow, D., Gondan, M., Agutter, J., Haar, M.,⋯Staggers, N. (2013). Evaluation of the effect of information integration in displays for ICU nurses on situation awareness and task completion time: A prospective randomized controlled study. *International Journal of Medical Informatics*, *82*, 665–675.

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica, 134,* 372–384.

Kristjánsson, Á., & Nakayama, K. (2003). A primitive memory system for the deployment of transient attention. *Perception & Psychophysics, 65,* 711–724.

Lansman, M., Farr, S., & Hunt, E. (1984). Expectancy and dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 195–204.

Liu, T., Fuller, S., & Carrasco, M. (2006). Attention alters the appearance of motion coherence. *Psychonomic Bulletin & Review, 13,* 1091–1096.

Lu, Z.-L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research, 38,* 1183–1198.

Luce, R. D. (1986). *Response times. Their role in inferring mental organization*. New York: Oxford University Press.

McDonald, J. J., Teder-Sälejärvi, W. A., Heraldez, D., & Hillyard, S. A. (2001). Electrophysiological evidence for the missing link in crossmodal attention. *Canadian Journal of Experimental Psychology, 55,* 141–149.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology, 14,* 247–279.

Miller, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics, 40,* 331–343.

Miller, J., Beutinger, D., & Ulrich, R. (2009). Visuospatial attention and redundancy gain. *Psychological Research, 73,* 254–262.

Mordkoff, J. T., & Danek, R. H. (2011). Dividing attention between color and shape revisited: redundant targets coactivate only when parts of the same perceptual object. *Attention, Perception, & Psychophysics, 73,* 103–112.

Mordkoff, T., & Yantis, S. (1993). Dividing attention between color and shape: Evidence for coactivation. *Perception & Psychophysics, 53,* 357–366.

Müller, H. J., & Rabbitt, P. M. A. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 315–330.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32,* 3–25.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X* (pp. 531–556). London: Lawrence Earlbaum Associates Ltd.

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. In H. L. Pick (Ed.), *Modes of perceiving and processing information* (pp. 137–157). New York, NY: Wiley.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109,* 160–174.

Posner, M. I., Inhoff, A. W., Friedrich, F. J., & Cohen, A. (1987). Isolating attentional systems: a cognitive-anatomical analysis. *Psychobiology*, *15*, 107–121.

Prime, D. J., McDonald, J. J., Green, J., & Ward, L. M. (2008). When cross-modal attention fails. *Canadian Journal of Experimental Psychology, 62,* 192–197.

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, *24*, 574–590.

Schröter, H., Ulrich, R., & Miller, J. (2007). Effects of redundant auditory stimuli on reaction time. *Psychonomic Bulletin & Review, 14,* 39–44.

Schwarz, W. (1994). Diffusion, superposition, and the redundant targets effect. *Journal of Mathematical Psychology, 38,* 504–520.

Schwarz, W. (2006). On the relationship between the redundant signals effect and temporal order judgments: parametric data and a new model. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 558–573.

Spence, C., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms facilitate sound

localization. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 555–574.

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1005–1030.

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics, 59,* 1–22.

Spence, C., McDonald, J., & Driver, J. (2004). Exogenous spatial cueing studies of human crossmodal attention and multisensory integration. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention*. New York, NY: Oxford University Press.

Talsma, D., & Kok, A. (2002). Intermodal spatial attention differs between vision and audition: an event-related potential analysis. *Psychophysiology, 39,* 689–706.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Science, 14,* 400–410.

Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience, 17,* 1098–1114.

Theeuewes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics, 49,* 83–90.

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

Vroomen, J., Bertelson, P., & de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, *63*, 651–659.

Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14,* 3–22.

Ward, L. M. (1994). Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. *Canadian Journal of Experimental Psychology, 48,* 242–259.

Ward, L. M., McDonald, J. J., & Lin, D. (2000). On asymmetries in cross-modal spatial attention orienting. *Perception & Psychophysics, 62,* 1258–1264.

Welch, R. B., & Warren, D. H. (1986). Intersensory interactions. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 25/1–25/36). New York: Wiley.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 601–621.

Zou, H., Müller, H. J., & Shi, Z. (2012). Non-spatial sounds regulate eye movements and enhance visual search. *Journal of Vision, 12,* 1–18.