Exogenous orienting of crossmodal attention in 3-D space: Support for a depth-aware crossmodal attentional system

**Nathan Van der Stoep · Tanja C. W. Nijboer ·
Stefan Van der Stigchel**

Abstract The aim of the present study was to investigate exogenous crossmodal orienting of attention in three-dimensional (3-D) space. Most studies in which the orienting of attention has been examined in 3-D space concerned either exogenous intramodal or endogenous crossmodal attention. Evidence for exogenous *crossmodal* orienting of attention in depth is lacking. Endogenous and exogenous attention are behaviorally different, suggesting that they are two different mechanisms. We used the orthogonal spatial-cueing paradigm and presented auditory exogenous cues at one of four possible locations in near or far space before the onset of a visual target. Cues could be presented at the same (valid) or at a different (invalid) depth from the target (radial validity), and on the same (valid) or on a different (invalid) side (horizontal validity), whereas we blocked the depth at which visual targets were presented. Next to an overall validity effect (valid RTs < invalid RTs) in horizontal space, we observed an interaction between the horizontal and radial validity of the cue: The horizontal validity effect was present only when the cue and the target were presented at the same depth. No horizontal validity effect was observed when the cue and the target were presented at different depths. These results suggest that exogenous crossmodal attention is "depth-aware," and they are discussed in the context of the supramodal hypothesis of attention.

Keywords Cueing · Exogenous · Orthogonal · Space · 3-D · Attention

In a typical exogenous crossmodal spatial-cueing study, a spatial cue is presented on the same side as the target or on the opposite side. Exogenous spatial cues presented in one modality are able to speed up responses to target stimuli presented in a different modality when they are presented at approximately the same location as the target, as compared to when they are presented at opposite locations (i.e., a validity effect). These crossmodal exogenous cueing effects have now been reported for all combinations of visual, auditory, and tactile stimuli (Spence & MacDonald, 2004), suggesting that exogenous spatial attention operates in a supramodal fashion (e.g., Eimer & Van Velzen, 2002; Farah, Wong, Monheit, & Morrow, 1989).

In most studies in which exogenous crossmodal orienting of attention has been examined, cueing effects were assessed in a single plane of depth. Yet, in real-life environments, visual and auditory sources can appear at various distances from the observer. Studies with healthy individuals in which exogenous intramodal attention was examined at different depths have indicated that exogenous visual cues are able to attract visual attention to a specific three-dimensional (3-D) location (Atchley, Kramer, Andersen, & Theeuwes, 1997; Bauer, Plinge, Ehrenstein, Rinkenauer, & Grosjean, 2011; Theeuwes & Pratt, 2003). Although it is known that exogenous *intramodal* attention operates in depth, evidence is scarce for exogenous *crossmodal* orienting of attention in depth. In a study by Ho and Spence (2005), the facilitating effects of exogenous and endogenous crossmodal

N. Van der Stoep (✉) · T. C. W. Nijboer · S. Van der Stigchel
Department of Experimental Psychology, Helmholtz Institute,
Utrecht University, Heidelberglaan 2, 3584 CS Utrecht,
The Netherlands
e-mail: n.vanderstoep@uu.nl

T. C. W. Nijboer
e-mail: T.C.W.Nijboer@uu.nl

S. Van der Stigchel
e-mail: S.VanderStigchel@uu.nl

T. C. W. Nijboer
Brain Center Rudolf Magnus and Center of Excellence for
Rehabilitation Medicine, University Medical Center Utrecht and
De Hoogstraat Rehabilitation, Utrecht, The Netherlands

attention were investigated in a simulated driving setup with cues and targets presented in front of and behind the participant. Responses were faster when the cue and the target originated from the same location in space. Although this indicates that attention can be crossmodally attracted to a location in 3-D space either in front of or behind a driver, it does not show whether exogenous crossmodal attention can be shifted in different planes of depth in front of the participant. In the one study that did look into *crossmodal* orienting of attention to different frontal depth planes, *endogenous* crossmodal attention was manipulated (Couyoumdjian, Di Nocera, & Ferlazzo, 2003; see Downing & Pinker, 1985, for intramodal endogenous attention in frontal depth). Although they recruit overlapping brain networks (Kim et al., 1999), exogenous and endogenous orienting of attention appear to have different properties (e.g., they have different time courses and are differently affected by cognitive load; see Berger, Henik, & Rafal, 2005). These differences indicate that exogenous and endogenous crossmodal forms of attention do not necessarily work in the same way, and may interact under, for example, high task demands.

So far, previous studies have thus provided support for the idea that *intramodal* exogenous and *intra*- and *crossmodal endogenous* attention are able to shift in depth. However, it is currently unclear whether and how *crossmodal exogenous* attention can be shifted in frontal space. In order to investigate this, we used the orthogonal spatial-cueing paradigm (Spence & Driver, 1997) and presented auditory cues and visual targets in either a near or a far depth plane. We blocked the plane of depth at which targets were presented, because otherwise near-space targets would occlude far-space targets, and we randomized the depth at which cues were presented. If exogenous crossmodal attention is "depth-aware," we expected to find a validity effect in the horizontal dimension (i.e., a classic validity effect) when cues were presented at the same depth as the target, but not when the cue and target were presented at different depths. In contrast, if exogenous crossmodal attention is not "depth-aware," the validity effects should not differ for cues presented at the same depth as or at different depths from the target.

Materials and method

Participants

On the basis of previous studies on attention in 3-D space (Atchley et al., 1997; Couyoumdjian et al., 2003; Theeuwes & Pratt, 2003), in which the samples varied between 10 and 24 participants, we included 16 healthy participants (13 female, three male; mean age = 22.44 years, SD = 1.90) who received course credits for their participation. All participants reported

normal or corrected-to-normal visual acuity and no hearing problems, and showed normal performance on a short left–right sound localization task (see below). The experiment was performed in accordance with the Declaration of Helsinki, and participants signed informed consent before the start of the experiment.

Apparatus

To project the visual stimuli on a black canvas (near, 75 × 60 cm; far, 170 × 170 cm), we used a Toshiba TLP-T621 LCD projector (60 Hz). Four speakers (Harman/Kardon HK206, frequency response: 90–20000 Hz) were used to present the auditory cues. A chinrest was used to stabilize the participant's head and to keep the distance between the participant and the projection largely stable across participants.

Stimuli, task, and procedure

The loudness, and even more so the direct-to-reverberant energy ratio of a sound, provides information about the distance of a sound source and enables us to estimate its approximate distance in closed environments (Bronkhorst & Houtgast, 1999). In order to investigate the influence of exogenous auditory cues from different locations in 3-D space on visual information processing, the cues in our experiment should vary on both properties, depending on the distance between the auditory source and the observer. This would ensure that the brain received enough information to estimate the approximate location of the sound source, and possibly to attract attention to that location. Auditory cues consisted of a 75-ms, 2000-Hz tone (10-ms rise and fall of the signal) of 100 dB(A) SPL, as measured with an audiometer directly in front of the speakers. We used a sine wave as the auditory cue, to ensure that left and right could be distinguished, but not elevation (up and down) (Frens & Van Opstal, 1995). This was important, because participants had to indicate whether a visual target was presented above or below the vertical center of the screen in the main experiment. The sine waveform ensured that the auditory cue could not be used as a landmark for visual target localization. Auditory cues that were presented in far space had a lower SPL than did auditory cues that were presented in near space, as measured with an audiometer from the distance at which the ears of the participant were located during the experiment [near space ±90 dB(A) SPL, far space ±80 dB(A) SPL]. In addition to objective measurements of SPL and inspection of the direct-to-reverberant profile of the auditory cues, we also behaviorally confirmed

these properties in a pilot study in which we examined whether the 3-D localization performance of the auditory cues was above chance.[1]

Before the start of the main experiment, each participant performed a short sound localization task to verify that he or she was able to hear whether auditory cues were presented on the left or the right side of the room. The sound localization task consisted of 20 presentations of the auditory cue from a random speaker (five presentations from each of the four locations). Participants had to maintain fixation on a fixation cross ($0.5° \times 0.5°$, 0.20 cd/m^2 as measured with a PhotoResearch SpectraScan PR 650 spectrometer) presented on a light gray background (4.79 cd/m^2) in the center of a screen in near space (at 80 cm distance). The participants were instructed to indicate whether the sound was coming from the left or the right side of the room by using two buttons. All participants performed the hearing task with above-chance accuracy.

As in the hearing task, participants in the main experiment were tested in a darkened room, with only the light of the projector illuminating the room. Visual targets were projected in either near (80 cm) or far (220 cm) space and were corrected for visual angle. Presenting stimuli at the same visual angle in different planes of depth has also been done in other studies in which attention was investigated in depth with rather large distances between depth planes (e.g., Couyoumdjian et al., 2003; Downing & Pinker, 1985). In the near-space condition, both screens were present, but targets were only projected on the near-space screen. In the far-space condition, the near-space screen was removed from the setup (target sizes were corrected for visual angle digitally). Speakers were positioned at four locations: near-left, near-right, far-left, and far-right space. Schematic top views of the experimental setup are shown in Fig. 1. Each block, visual targets were presented in one space only (near or far space), whereas auditory cues were randomly presented from one of the four speakers located in near or far space. The order of the regions of space in which visual targets were presented first was blocked and counterbalanced across participants. Each speaker was placed outside the light of the projection of the beamer on the left and right sides of the screen. As a result, the speakers in near space were located 23° from the fixation cross and 19° in far space. The experiment started with 20 practice trials for the participants to get used to the task.

Participants were instructed to gaze at a gray fixation cross (size: $0.91° \times 0.91°$ in near and in far space, 0.38 cd/m^2 in near

space) presented on a black background (<0.15 cd/m^2 in near and far space) in the center of the screen at a height of 34 cm above the table.[2] After 1,000 ms, an auditory cue was presented at one of the four locations while the fixation cross remained on the screen. Auditory cues could be presented on the same side as (valid) or on the opposite side from (invalid) the visual target, and at the same (valid) or at a different (invalid) depth. The procedure of the experiment is shown in Fig. 2.

Also, in a no-cue condition, no cue was presented before the target appeared. Targets were presented in two regions of space, resulting in ten conditions. Each condition contained 80 trials, adding up to a total of 800 trials. A break was provided after 200 trials during a block, and participants could press the space bar to continue. After 400 trials, the visual targets were presented in the other region of space (first far and then near, or vice versa), and another 400 trials were presented. The stimulus onset asynchrony varied between 90 and 250 ms and was always followed by the presentation of a target. The targets were filled gray circles with a diameter of 2.60° in either near space (0.38 cd/m^2) or far space. The target location was randomized and could be presented either above or below the vertical center of the screen, to the left or the right of the fixation cross. The horizontal distance from the fixation cross to the target was 14.16°, and the vertical distance from the fixation cross to the middle of the target (either above or below) was 3.9°. The target disappeared upon response. Participants were instructed to press the number-pad "5" key for an upper target, and the number-pad "2" key for a lower target. The maximum response duration was set to 2,000 ms, after which the target disappeared automatically. The intertrial interval consisted of the presentation of the background alone, with a duration of 1,200 ms.

Data analysis

Preprocessing Practice trials were excluded from both the accuracy and the response time (RT) analyses. We only analyzed the RTs of correct trials. In addition, trials on which the RT was below 100 ms or above 1,000 ms were removed from further analysis, since they were considered to be the results of anticipation or of not attending to the experiment, respectively. RTs were regarded as outliers when they exceeded two-and-a-half standard deviations above or below the group mean of a condition. On average, 6 % of the trials were removed from further analysis when targets were presented in near space, and 5 % of the trials were removed when targets were presented in far space.

---

[1] Ten participants who were not part of the main experiment sample (five female, five male; mean age = 25.70 years, $SD = 3.13$) were tested with a four-choice localization task in the same setup as in the main experiment. They were instructed to maintain fixation on a black fixation cross ($0.5° \times 0.5°$, 0.20 cd/m^2) presented on a light gray background (4.79 cd/m^2) in near space (80 cm). The average accuracy of 3-D localization was significantly above chance ($M = .55$, $SD = .16$) [$t(9) = -6.00$, $p < .001$].

---

[2] Note that we did not track eye movements, and therefore that we cannot draw any conclusions about whether any exogenous cueing effects were due to covert or overt exogenous orienting of attention. The results will still be able to indicate whether attention can be exogenously shifted in 3-D space in a crossmodal way.
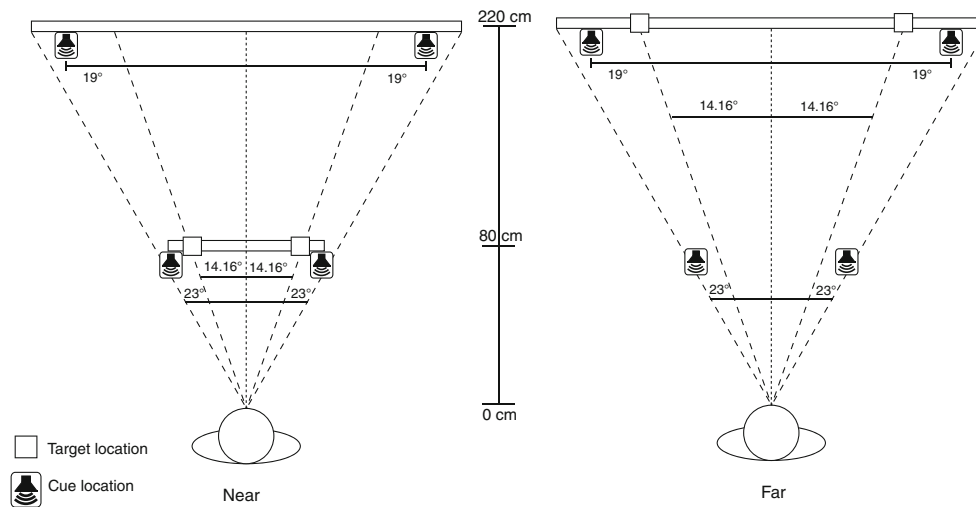
Fig. 1 Schematic top views of the experimental setup in the near-space (left panel) and far-space (right panel) conditions

Statistical analysis First, to investigate whether the overall effect of horizontal cue type (i.e., a "2-D" cueing effect), we performed a repeated measures analysis of variance (ANOVA) with the within-subjects factor Horizontal Cue Type (no cue, valid cue, invalid cue), with RTs being averaged over radial cue validities and distances of the target.

Second, to more closely investigate how the distance between the cue and the target in the radial plane influenced RTs, we performed a 2 × 2 × 2 repeated measures ANOVA on RTs and accuracy, with the factors Target Space (near, far), Horizontal Cue Validity (valid, invalid), and Radial Cue Validity (valid, invalid). Note that we did not include the no-cue condition in the design, because the no-cue condition could not be valid or invalid in the horizontal and radial dimensions. Paired-samples *t* tests were done to compare differences between each of the conditions, and the resulting *p* values were Bonferroni corrected where applicable.
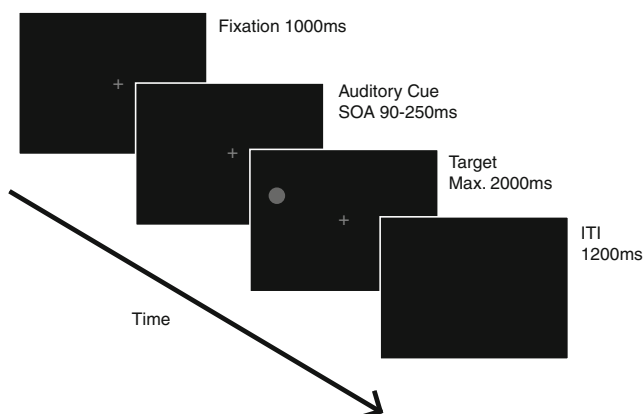
Fig. 2 Schematic representation of the procedure of the experiment

Results and discussion

Left–right sound localization task

None of the participants were excluded from the analysis on the basis of the short left–right hearing task, and the group as a whole performed well above chance (mean accuracy = .97, *SE* = .01), $t(15) = 42.37$, $p < .001$. All participants performed the task with an accuracy of at least .85.

Accuracy

The repeated measures ANOVA with the factors Target Space (near, far), Horizontal Cue Validity, and Radial Cue Validity revealed no significant main effects or interactions (all *p*s > .05). The average accuracy was .976 (*SE* = .005). All participants performed with an accuracy of at least .91.

Response times

The results of the repeated measures ANOVA indicated a significant effect of cue type [valid, invalid, or no cue: $F(1.120, 16.799) = 109.946$, $p < .001$, Greenhouse–Geisser epsilon = .560, $\eta^2 = .880$]. Pairwise comparisons indicated that RTs were significantly slower in the no-cue condition (*M* = 521 ms, *SE* = 15) than in the invalid condition (470 ms, *SE* = 13, $p < .001$) and the valid condition (462 ms, *SE* = 13, $p < .001$). In addition, RTs were significantly faster on valid than on invalid trials ($p = .001$). This indicated that both types of auditory cues facilitated RTs to the target, relative to when no auditory cue was presented (i.e., an alerting effect), with valid cues resulting in the fastest responses.

The analysis with the factors Target Space, Horizontal Cue Validity, and Radial Cue Validity revealed a significant main effect of horizontal cue validity [$F(1, 15) = 24.374$, $p < .001$,

$\eta^2 = .619$]. RTs on horizontally validly cued trials were significantly faster than those on horizontally invalidly cued trials (462 ms, $SE = 13$, vs. 470 ms, $SE = 13$). We did not find a main effect of target space [$F(1, 15) = 1.587$, $p = .227$, $\eta^2 = .096$] and of radial cue validity [$F(1, 15) = 0.055$, $p = .818$, $\eta^2 = .004$]. The interaction between horizontal cue validity and radial cue validity was significant [$F(1, 15) = 6.390$, $p = .023$, $\eta^2 = .299$]. The magnitude of the horizontal validity effect depended on whether the radial distances between the cue and the target were the same. We collapsed the near- and far-space conditions in subsequent analyses, because of the lack of a main effect of target space. Figure 3c shows the mean RT in each condition. The difference between the horizontal valid (461 ms, $SE = 13$) and horizontal invalid (472 ms, $SE = 14$) conditions was 11 ms when the cue was presented at the same depth as the target (radial valid). When the cue and the target were presented at different depths (radial invalid), the difference between the horizontal valid (464 ms, $SE = 14$) and horizontal invalid (467 ms, $SE = 13$) conditions was 3 ms. None of the other interactions were significant (all ps > .1).

Planned comparisons revealed a validity effect (horizontal invalid– horizontal valid) when the cue and target were presented at the same distance [mean horizontal validity effect = 11 ms, SE of the difference = 2 ms; $t(15) = -5.360$, $p < .001$], but not when they were presented at different distances [mean horizontal validity effect = 3 ms, SE of the difference = 2 ms; $t(15) = -1.359$, $p = .350$]. The size of the validity effect was significantly larger when the cue and target were presented at the same distance (11 ms), as compared to

when they were presented a different distances (3 ms) [$t(15) = 2.528$, $p = .023$]. The difference between radial valid and radial invalid cues was not significantly different when cues were horizontally valid (mean radial validity effect = 4 ms) [$t(15) = -1.744$, $p = .194$], nor when cues were horizontally invalid (mean radial validity effect = −4 ms) [$t(15) = 2.092$, $p = .105$].

Discussion

The aim of the present study was to investigate the nature of exogenous crossmodal attention in 3-D space. We presented visual targets in either near or far space, and exogenous auditory cues from one of four locations: near left, near right, far left, or far right. The results indicated an overall cue validity effect in the horizontal dimension. More interestingly, the presence of a horizontal validity effect was dependent on whether the cue and the target were presented at the same depth: A validity effect was only present when the cue and the target were presented at the same depth, and not when the cue and the target were presented at different depths. These findings suggest that exogenous crossmodal attention is "depth-aware." In contrast, if exogenous crossmodal attention were not "depth-aware," we should have observed a horizontal validity effect when the cue and the target were presented both at the same and at different depths from the target. This was not the case.

In our study, the horizontal validity effect for cues and targets presented at the same depth did not differ between
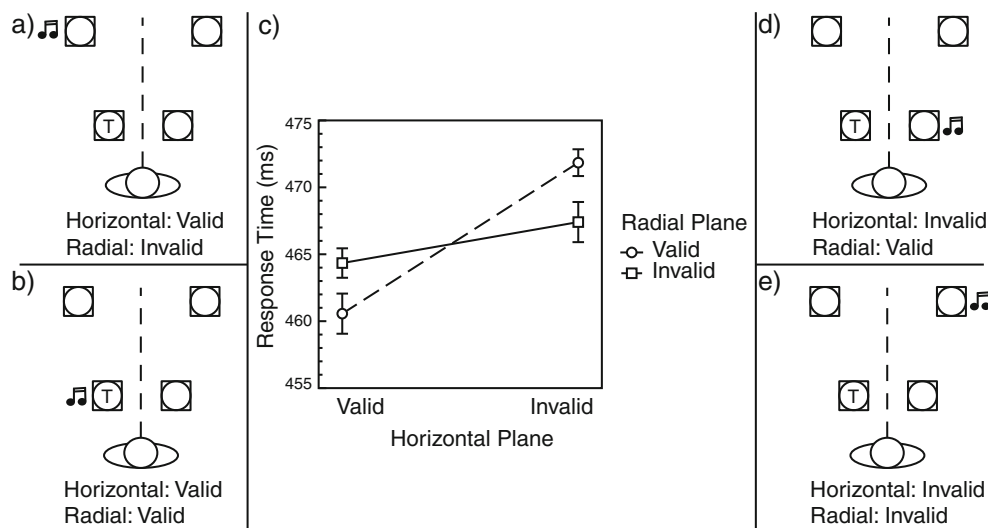
Fig. 3 Schematic representations of the different cue conditions (with their corresponding data points depicted in panel c) are shown in panels a, b, d, and e. The notes represent auditory cue locations, and the "T"s indicate target locations (target distance was blocked). Panel c depicts the mean response times for valid and invalid trials in the horizontal and radial planes. Error bars represent standard errors of the means, without between-subjects variability for graphical purposes (Cousineau, 2005)

visual targets presented in near and far space. This finding is in contrast with those from other studies on attentional orienting in depth, in which an asymmetry was observed between the cueing effects for targets in near and targets in far space (e.g., Chen, Weidner, Vossel, Weiss, & Fink, 2012; Downing & Pinker, 1985). In these studies, in which endogenous attention was manipulated, participants were faster to respond to targets that were presented between the participant and the focus of *endogenous* attention, as compared to targets that were presented beyond the focus of *endogenous* attention. The lack of an asymmetry in attentional reorienting in depth in the present study may be explained by the fact that we blocked target distance: In our study, participants had no need to attend to multiple planes of depth, which might have caused an endogenous focus on one depth plane (i.e., 100 % endogenous validity), possibly overruling any asymmetry of attention in depth.

The conclusion that exogenous crossmodal attention is "depth-aware" seems to be in contrast with the results of the short four-choice localization task. Although participants could localize the cue significantly above chance, their accuracy was rather low (mean = 55 %). Still, the depth of the cues in the main experiment influenced the presence of the horizontal validity effect. This suggests that despite participants' being poor at consciously locating this type of sound (a sine wave), the brain did process the depth information of auditory sources.

Our findings are in line with the results from studies in which exogenous *intramodal* orienting of attention was investigated in 3-D space (Atchley et al., 1997; Theeuwes & Pratt, 2003). For example, Theeuwes and Pratt also found that the validity effect was stronger when the cue and target were presented at the same distance from the observer. Here, however, we extended these findings by showing that the *crossmodal* exogenous cues were also able to automatically attract attention to different depth planes in "real" 3-D space. Altogether, the previous and present results therefore seem to fit with the theory of a supramodal attentional system (Eimer & Van Velzen, 2002; Farah et al., 1989; Macaluso, Frith, & Driver, 2002) that processes spatial information from the auditory and visual modalities, despite differences in spatial reference frame (retinotopic [Gardner, Merriam, Movshon, & Heeger, 2008] vs. head-centered [Andersen, 1997]). A candidate region for supporting such a supramodal attentional system is the posterior parietal cortex, and more specifically the multisensory lateral intraparietal area (area LIP; Andersen, 1997). Crossmodal interactions also seem to depend on feedforward and feedback connections between unimodal and multisensory areas (Macaluso & Driver, 2005).

Taken together, our results indicate that the *exogenous* orienting of crossmodal attention is "depth-aware," and they contribute to the further understanding of crossmodal interactions in 3-D space.

References

Andersen, R. A. (1997). Multimodal integration for the representation of space in the posterior parietal cortex. *Philosophical Transactions of the Royal Society B, 352,* 1421–1428.

Atchley, P., Kramer, A. F., Andersen, G. J., & Theeuwes, J. (1997). Spatial cuing in a stereoscopic display: Evidence for a "depth-aware" attentional focus. *Psychonomic Bulletin & Review, 4,* 524–529. doi:10.3758/BF03214343

Bauer, D., Plinge, A., Ehrenstein, W. H., Rinkenauer, G., & Grosjean, M. (2011). Spatial orienting of attention in stereo depth. *Psychological Research, 76,* 730–735.

Berger, A., Henik, A., & Rafal, R. (2005). Competition between endogenous and exogenous orienting of visual attention. *Journal of Experimental Psychology: General, 134,* 207–221. doi:10.1037/0096-3445.134.2.207

Bronkhorst, A. W., & Houtgast, T. (1999). Auditory distance perception in rooms. *Nature, 397,* 517–520.

Chen, Q., Weidner, R., Vossel, S., Weiss, P. H., & Fink, G. R. (2012). Neural mechanisms of attentional reorienting in three-dimensional space. *Journal of Neuroscience, 32,* 13352–13362.

Cousineau, D. (2005). Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1,* 42–45.

Couyoumdjian, A., Di Nocera, F., & Ferlazzo, F. (2003). Functional representation of 3D space in endogenous attention shifts. *Quarterly Journal of Experimental Psychology, 56A,* 155–183.

Downing, C. J., & Pinker, S. (1985). The spatial structure of visual attention. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 171–187). Hillsdale, NJ: Erlbaum.

Eimer, M., & Van Velzen, J. (2002). Crossmodal links in spatial attention are mediated by supramodal control processes: evidence from event-related potentials. *Psychophysiology, 39,* 437–449.

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

Frens, M. A., & Van Opstal, A. J. (1995). A quantitative study of auditory-evoked saccadic eye movements in two dimensions. *Experimental Brain Research, 107,* 103–117.

Gardner, J. L., Merriam, E. P., Movshon, J. A., & Heeger, D. J. (2008). Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. *Journal of Neuroscience, 28,* 3988–3999.

Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of Experimental Psychology: Applied, 11,* 157–174.

Kim, Y., Gitelman, D. R., Nobre, A. C., Parrish, T. B., LaBar, K. S., & Mesulam, M. M. (1999). The large-scale neural network for spatial attention displays multifunctional overlap but differential asymmetry. *NeuroImage, 9,* 269–277.

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: A window onto functional integration in the human brain. *Trends in Neurosciences, 28,* 264–271.

Macaluso, E., Frith, C. D., & Driver, J. (2002). Supramodal effects of cover spatial orienting triggered by visual or tactile events. *Journal of Cognitive Neuroscience, 14,* 389–401.

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics, 59,* 1–22. doi:10.3758/BF03206843

Spence, C., & MacDonald, J. (2004). The cross-modal consequences of the exogenous spatial orienting of attention. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (1st ed., pp. 3–26). Cambridge, MA: MIT Press.

Theeuwes, J., & Pratt, J. (2003). Inhibition of return spreads across 3-D space. *Psychonomic Bulletin & Review, 10,* 616–620. doi:10.3758/BF03196523