Available online at www.sciencedirect.com
SCIENCE DIRECT
COGNITION
Cognition 96 (2005) B55–B64
www.elsevier.com/locate/COGNIT

Brief article

Crossmodal object-based attention: Auditory objects affect visual processing

Massimo Turatto[a,*], Veronica Mazza[b], Carlo Umiltà[b]

[a]Department of Cognitive Science and Education, University of Trento, Via Matteo del Ben, 5, 38068 Rovereto (TN), Italy
[b]University of Padova, Padova, Italy

Received 28 May 2004; accepted 6 December 2004

Abstract

According to the object-based view, visual attention can be deployed to "objects" or perceptual units, regardless of spatial locations. Recently, however, the notion of object has also been extended to the auditory domain, with some authors suggesting possible interactions between visual and auditory objects. Here we show that task-irrelevant auditory objects may affect the deployment of visual attention, providing evidence that crossmodal links can also occur at an object-based level. Hence, in addition to the well documented control of visual objects over what we hear, our findings demonstrate that, in some cases, auditory objects can affect visual processing.

Keywords: Crossmodal attention; Object-based attention; Auditory objects; Sensory modalities

Attention is an important cognitive function by means of which the human cognitive system is able to select the information relevant for the current behaviour. According to the "space-based" view (Posner, Snyder, & Davidson, 1980; for a review see Cave & Bichot, 1999), selection takes place in spatial coordinates, with the spotlight of attention moving in different regions of the visual field. Alternatively, following Neisser's (1967) original suggestion, Duncan (1984) proposed the "object-based" view, claiming that the units over which attention processes operate are discrete visual objects (for a review see Scholl, 2001; also see Driver, Davis, Russell, Turatto, & Freeman, 2001).

* Corresponding author. Tel.: +39 0464 483529.
E-mail address: massimo.turatto@unitn.it (M. Turatto).

doi:10.1016/j.cognition.2004.12.001

Egly, Driver, and Rafal (1994; also see Moore, Yantis, & Vaughan, 1998) reconciled these two apparently opposite theoretical positions, providing evidence for both a space-based and an object-based component of visual attention. In their study, participants were presented with a display consisting of two outlined rectangles placed on either side of the fixation point (arranged either vertically or horizontally). The authors cued the observers to the end of one rectangle. After a 75%-valid cue, the target (a luminance increment) occurred at an end of one of the two objects, and the observers had to detect it as quickly as possible. Two types of invalid trials were possible. On "within-object" invalid trials, the target appeared at the uncued end of the cued object, whereas on "between-objects" invalid trials the target appeared at the end corresponding to the cued one but in the uncued object. Results showed that response times (RTs) were lower on valid trials than on invalid trials, in accordance with the space-based component of attention. However, RTs were lower on invalid within-object than on invalid between-objects trials. Because the spatial distance between the cue and the target location was identical in the two invalid conditions, the lower RTs on the invalid within-object trials were evidence of a "same-object advantage", showing an object-based component of attention (Egly et al., 1994).

In those same years, a new research effort was concentrated on the relation between attention and sensory modalities (e.g. Spence & Driver, 1996; Ward, 1994). Different issues, previously neglected, gained interest within this multi-sensory attentional perspective. For example, an important question to address was how the brain organizes inputs from senses having different spatial organisation (audition is first tonotopic then head-centred; touch is somatotopic; vision is retinotopic). Crucial for the aim of the present study was the issue of whether attention could be directed in space across different modalities. As a general method, crossmodal links in spatial attention have been studied, and documented, by presenting a cue in one modality and the target in another modality (Spence & Driver, 1996; Ward, 1994). Typically, results show that, for both exogenous and endogenous orienting, a spatially localized stimulus in a given modality usually elicits a shift of covert attention also in other modalities (Driver & Spence, 2004; McDonald, Teder-Sälejärvi, Heraldez, & Hillyard, 2001; Spence, Pavani, & Driver, 2000; Ward, McDonald, & Lin, 2000).

Crossmodal attention studies to date seem to have investigated only spatial links between modalities (but see Turatto, Benso, Galfano, & Umiltà, 2002; Turatto, Galfano, Bridgeman, & Umiltà, 2004), without exploring whether such crossmodal links might also take place at an "object-based" level. The lack of studies exploring this possibility is likely due to the implicit assumption that only visible units can be referred to as "objects". However, in accordance with Bregman (1990) and with Kubovy and Van Vanlkenburg (2001), it would seem that the notion of object can be extended to audition, and possibly to touch. Here we provide evidence that auditory objects can affect the deployment of visual attention, thus revealing possible crossmodal links at an object-based level.

1. Experiment 1

We adapted the original experimental paradigm devised by Egly et al. (1994) to create two auditory objects, separated along the horizontal axis. They were obtained by playing a tone from the left pair of loudspeakers, one above and one below fixation, and a different tone from the right pair. Hence, the two objects were perceived as being located at each side of fixation. Kubovy and Van Valkenburg (2001; also see Bregman, 1990) have maintained that for perceiving two auditory objects it is not necessary to have two auditory sources originating from separate spatial locations, as, in their view, frequency separation is a sufficient condition to produce distinct auditory objects. Nevertheless, to enhance the perception of distinct auditory objects, in the present experiment we presented two streams of tones of different frequency from different spatial locations (also see Wightman & Jenison, 1995).

Importantly, the auditory objects we presented were completely task irrelevant. Indeed, although the visual target appeared in a spatial position occupied by one of the two auditory objects, the most likely spatial position for the target to appear was always signalled by a visual cue. In addition, participants were asked to perform a target colour discrimination task, which was clearly unrelated to the auditory information.

Our hypothesis was that the presence of *two different auditory perceptual units differing in spatial location* would affect the distribution of visual attention. Specifically, we hypothesised that when attention was cued to a given spatial position, it was also automatically allocated to the corresponding auditory object on the same side. Hence, on the basis of Egly et al.'s (1994) results, we expected a faster visual target discrimination on invalid trials in which the target occurred in the same auditory object as compared to invalid trials in which the target occurred in a different auditory object.

1.1. Method

1.1.1. Participants

Twelve undergraduate students from the University of Padova, all right-handed, participated in the experiment. All had normal or corrected-to-normal vision and hearing, and were not aware of the purpose of the experiment.

1.1.2. Apparatus and stimuli

The experiment was conducted in a dark (about 0.1 cd/m^2), sound-attenuated room. The auditory stimuli were provided by four loudspeakers, whose centres were positioned at the vertices of an imaginery square of 90 cm of side (see Fig. 1, panel A). The two auditory objects were obtained by having the left and right loudspeakers playing a train of 1000-Hz tones and 2000-Hz tones, respectively. The intensity of both tones was approximately 70 dB, and each tone was presented simultaneously to the high and low speakers on the appropriate side. Tones were gated on and off with two raised cosine ramps of 10 ms of duration. Each train consisted of a sequence of 10 tones lasting 100 ms, followed by a silence interval of 200 ms. In order to facilitate the spatial separation of the two auditory objects, the left and the right trains of tones were delivered asynchronously, with a temporal gap of 150 ms (see Fig. 1, panel B). Phenomenally, each auditory stream was heard as a train of sounds emanating from some region between the two vertical loudspeakers.[1]

[1] A few new independent subjects were interviewed and asked to describe of how they perceived the auditory streams. While some described the sound as emanating from an intermediate position between the two vertically arranged loudspeakers, others were less confident and accurate in indicating a precise point, but instead indicated a roughly defined region between the loudspeakers.

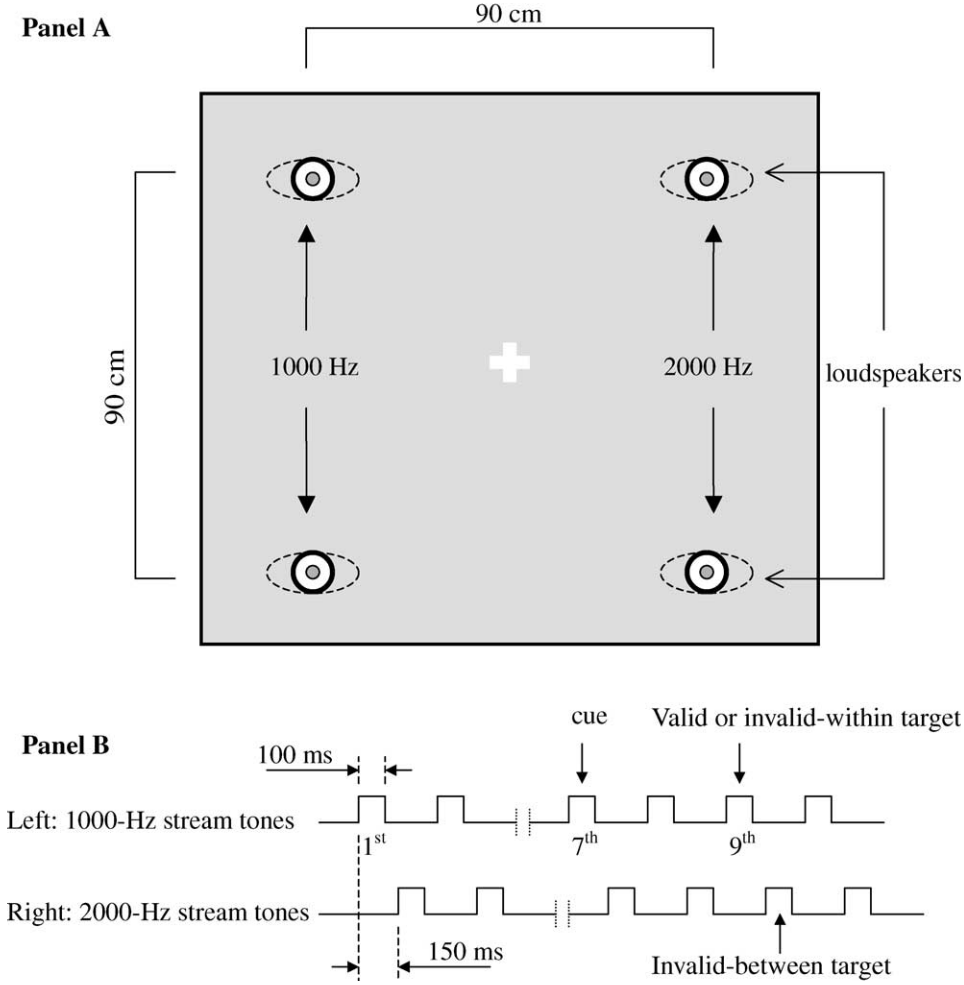

Fig. 1. Panel A depicts the experimental device used in Experiments 1 and 2. The four loudspeakers were placed behind a cardboard box. The pair of loudspeakers to the left played a stream of 1000-Hz tones, whereas the pair to the right played a stream of 2000-Hz tones. Panel B provides the temporal diagram of the two streams of tones in Experiment 1. The cue appeared always synchronized with the 7th tone of a given sequence, whereas the target was delivered always synchronized with either the 9th tone of the same auditory stream (valid or invalid within-object trials) or the 9th tone of the opposite auditory stream (invalid between-objects trials).

Importantly, no visible objects were present in the observer's visual field, as a black cardboard box was placed in front of the loudspeakers, rendering them invisible. The visual stimuli were provided by four trichromatic light emitting diodes (LEDs), each placed at the centre of the corresponding loudspeaker. The LEDs emerged from the cardboard box through four holes of about 3 cm of radius, which also allowed the cone of the loudspeakers to remain uncovered. In addition, a white cross arranged at the centre of the imaginary square served as fixation point. The cue consisted of an orange LED presented for 100 ms, whereas the target, lasting 100 ms, was randomly red or green. The cue and the target occurred equally likely in each of the four positions, and their onset was *always* synchronized with the 7th and 9th tone in the appropriate (left or right) auditory sequence.

1.1.3. Design and procedure

The experimental session consisted of 288 trials divided into two blocks, and a block of 40 practice trials. Each experimental block contained 96 valid trials, 24 invalid within-object trials, and 24 invalid between-objects trials. At the beginning of each trial, the trains of left and right auditory stimuli were delivered from the corresponding pair of loudspeakers. The visual cue was synchronized with the 7th tone in the left or right sequence, and was followed 2 tones later by the target. The inter trial interval was 1000 ms. On 66% of trials the cue and the target appeared in the same spatial position (valid trials). On 17% of trials the cue and the target appeared in two separate positions, which played the same tone (invalid within-object trials). On the remaining 17% of trials, the cue and the target appeared in two separate positions playing different tones (invalid between-objects trials). Participants were instructed to press, as quickly as possible, the “B” key when the target was red, and the “N” key when the target was green.

1.2. Results

Outliers, defined as RTs lower than 150 ms, were trimmed before data analysis (less than 1%). RTs were entered into a repeated measures analysis of variance (ANOVA), in which the factor was Type of trials (three levels: valid, invalid within-object, and invalid between-objects). Greenhouse-Geisser correction was used when appropriate. The factor was significant, $F(2,22)=17.106$, $P<0.001$. Planned comparisons (*t*-tests) revealed that participants were faster on valid trials ($M=530$; SD=75) than on either invalid within-object trials ($M=556$; SD=87), $t(11)=4.3$, $P<0.001$, or invalid between-objects trials ($M=573$; SD=94), $t(11)=4.86$, $P<0.001$. These results confirmed the effectiveness of the visual cue in directing attention, as predicted by the space-based view (e.g. Posner et al., 1980). More interestingly, the comparison between the two types of invalid trials revealed that RTs on invalid within-object trials were lower than RTs on invalid between-objects trials, $t(11)=2.44$, $P<0.016$ (see Fig. 2). Following Egly et al. (1994), this result can be taken as evidence of an object-based allocation of attention. Importantly, the only perceptual units that might have mediated such an object-based effect were the auditory objects.

Errors were less than 6% in all conditions, and when analysed the effect of Type of trials was not significant ($F<1$).

1.3. Discussion

Our hypothesis was that the perception of auditory objects might influence the allocation of visual attention. The results of the present experiment are congruent with our prediction. Because on both invalid within-object trials and invalid between-objects trials spatial distance between cued location and target location was identical, the RT difference

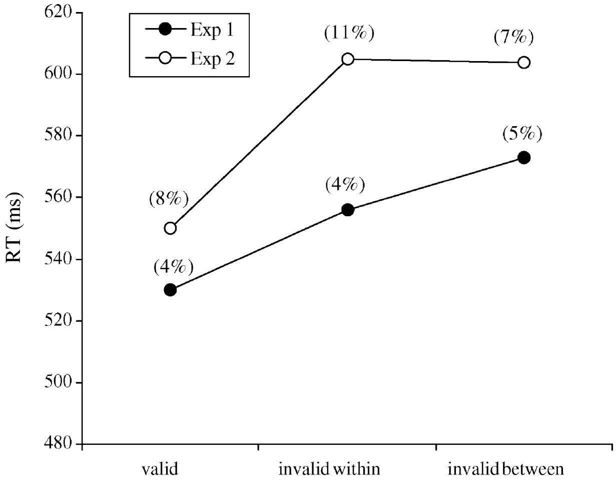

Fig. 2. Results from Experiments 1 and 2. In both experiments participants responded faster on valid trials than on invalid trials. However, target discrimination was faster on invalid within-object trials as compared to invalid between-objects trials only when left and right auditory streams were made of different tones, giving rise to distinct auditory objects (Experiment 1). When the tone frequency separation was removed, RTs for target discrimination did not differ between the two types of invalid trials (Experiment 2). Error rates are shown in parenthesis.

between these conditions is evidence of an object-based component of attention (also see Egly et al., 1994). Interestingly, as compared to Egly et al.'s study, in the present experiment there were no visual objects to which the observed object-based effect could be ascribed. Instead, the only perceptual units affecting the distribution of visual attention were the two auditory objects. This, in our view, demonstrates a crossmodal link in object-based attention. However, before accepting this conclusion, alternative explanations must be considered.[2]

2. Experiment 2

In Experiment 1 the cue and the target were always presented synchronously with the 7th and the 9th tone in the corresponding auditory stream on both invalid

[2] In Experiment 1 the reorienting of attention invariably occurred along the vertical meridian on invalid within-object trials, and along the horizontal meridian on invalid between-objects trials. In the literature, there is evidence that the movement of visual attention is slower when either the vertical or the horizontal meridians are crossed (e.g. Rizzolatti, Riggio, Dascola, & Umiltà, 1987). Therefore, we could not *a priori* exclude that the same-object advantage was due to a slower reorienting across the horizontal meridian as compared to the vertical one. To rule out this possibility, in an additional control experiment, not reported in full here, we replicated Experiment 1 without presenting any auditory object. Crucially, the RT difference between the two types of invalid trials disappeared (invalid within-object trials, $M=571$; invalid between-objects trials, $M=565$; $P>0.2$).

within-object and invalid between-objects trials. Because the two streams were delivered with a temporal gap of 150 ms, the resulting SOA between the cue and the target was 150 ms longer on invalid between-objects trials than on invalid within-object trials. We might have set the target occurrence at a fixed delay from the cue, but in this case the target would have been presented with a tone on invalid within-object trials and with no tones on invalid between-objects trials (see Fig. 1, panel B). Had this been the case, the RT advantage of invalid within-object trials over invalid between-objects trials could have been explained by the fact that perception is enhanced when the visual target temporally coincides with an auditory stimulus as compared to when it does not (Stein, London, Wilkinson, & Price, 1996; Vroomen & de Gelder, 2000). However, because the visual target was synchronized with a tone on both types of invalid trials, we could not exclude, as an alternative explanation, that the observed same-object advantage might be due to the SOA difference itself, rather than to the presence of two different auditory objects.

To address this issue, we carried out Experiment 2, which was identical to Experiment 1 but left and right tones had the same frequency (1000 Hz). Under this condition we can make two straightforward predictions. If the SOA difference between invalid-within object and invalid-between objects trials caused the RT pattern on invalid trials of Experiment 1, then we should observe the same result even when left and right tones are identical. By contrast, if our hypothesis that auditory objects modulated visual attention was correct, the RT advantage of same-object trials over different-objects trials should disappear when the frequency is identical for both left and right tones, as frequency separation is a necessary condition to produce distinct auditory objects.

2.1. Method

2.1.1. Participants

Twelve undergraduate students from the University of Padova, all right-handed, participated in this experiment. All had normal or corrected-to-normal vision and hearing, and were not aware of the purpose of the experiment.

2.1.2. Apparatus, stimuli, design and procedure

As in Experiment 1, but the right and left tones were set to 1000 Hz.

2.2. Results

Outliers were defined as previously (less than 1%). RTs were entered into a repeated measures ANOVA with Type of trials as a factor. Greenhouse-Geisser correction was used when appropriate. The factor was significant, $F(2,22)=10.553$, $P<0.001$. As in Experiment 1, planned comparisons revealed that participants were faster on valid trials ($M=550$, SD=96) than on either invalid within-object trials ($M=605$; SD=105), $t(11)=3.16$, $P<0.009$, or invalid between-objects trials ($M=604$; SD=98), $t(11)=4.13$,

$P<0.002$. Crucially, however, RTs on the two type of invalid trials were not significantly different[3], $t(11)=0.47$, $P>0.9$ (see Fig. 2).

Errors were less than 12% in all conditions, and when analysed the factor Type of trials was not significant ($F=1.05$).

2.3. Discussion

The crucial result of Experiment 1 might have been attributed to the longer SOA on invalid between-objects trials as compared to invalid within-object trials. By contrast, according to the object-based account, the results would be explained by the fact that different right and left tones created two distinct auditory objects. Here we showed that the same-object advantage emerged in Experiment 1 disappeared when the target occurred, at different SOAs, either in the left or right auditory stream of the same tone. This result is not consistent with the SOA hypothesis, whereas it is in accordance with the two auditory objects hypothesis. Accordingly, when the tone frequency separation was removed, left and right tones no longer created separate objects, which in turn caused the same-object advantage to vanish. This result also supports the idea that the effect observed in Experiment 1 did not depend on the mere presentation of auditory stimuli at visual target locations.

3. General discussion

In the last decade, crossmodal links in spatial attention between vision and audition have been well documented (Ward et al., 2000). However, spatial locations are not the only features over which attention can operate. Indeed, robust evidence exists that, under certain circumstances, attention is deployed to objects rather than to spatial locations (Scholl, 2001). In addition, many researchers have proposed that, despite our bias in considering only visible elements as objects, the notion of "objecthood" could also apply to the auditory modality (e.g. Bregman, 1990). The present study was aimed at addressing whether attentional crossmodal links between vision and audition could also be established at an object-based level, i.e. when auditory objects are provided during a visual task. The results showed that two task-irrelevant auditory objects affected the distribution of visual attention, with faster discrimination when the cue and the target occurred in the same auditory object than when they occurred in two different auditory objects.

[3] To avoid problems in interpreting null effects and to make our conclusions more convincing, we decided to perform an analysis considering the experiments as a between-subjects factor. Because the RT pattern was virtually identical to that of the other control experiment (see footnote 2), and in order to increase the power of the statistical analysis, we combined data from these two experiments and considered them as a single experiment. In a two-way ANOVA we compared the RT difference between invalid within-object and invalid between-object trials in Experiment 1 and in the combination of the two control experiments. Results showed a significant interaction between type of invalid trials and experiment, $F(1,34)=4.400$, $P<0.043$, confirming that the same-object advantage was evident only in Experiment 1.

There are at least two possible ways in which the auditory objects may have affected visual attention. On the one hand, each stream of sounds might have been perceived as emanating from a region roughly centred at participants' ear level, with a certain spatial extension above and below the middle point between the loudspeakers on the same side. If this was the case, upon cue presentation visual attention could have been spread over the area occupied by the synchronous auditory object, thus speeding up visual target discrimination on invalid within-object trials as compared to invalid between-objects trials. Alternatively, however, it is likely that the two sounds might have been heard as emanating from the intermediate position (i.e. exactly from the middle position) between the two vertically arranged loudspeakers on the same side. In this case, the advantage found for the invalid-within object trials might be better explained as a consequence of a crossmodal perceptual grouping between the auditory object and the visual events occurring on the same side. Specifically, when the visual cue appeared, attention could have been directed toward both the visual event and the auditory object synchronously presented with it. When the target appeared synchronous with the auditory object previously attended and grouped with the cue, target discrimination was faster than when the target occurred on the other side and synchronous with the previously unattended auditory object.

While further studies will address this issue more in detail, the present study provides evidence for a modulation of visual attention by auditory perceptual units.

Acknowledgements

We wish to thank Brian Scholl and three anonymous reviewers for very helpful comments on an early draft of this manuscript. We also thank Martina Rossi for her help in data collection of Experiment 2. This study was supported by grants from MIUR to M. T. and C. U.

References

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Cave, K. R., & Bichot, N. (1999). Visuospatial attention: Beyond a spotlight model. *Psychonomic Bulletin and Review*, *6*, 204–233.

Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. (2001). Segmentation, attention, and phenomenal visual objects. *Cognition*, *80*, 61–95.

Driver, J., & Spence, C. (2004). Crossmodal spatial attention: Evidence from human performance. In C. Spenc, & J. Driver (Eds.), *Crossmodal space and crossmodal attention*. Oxford: Oxford University Press.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*, 501–507.

Egly, R., Driver, J., & Rafal, R. (1994). Shifting visual attention between objects and locations: Evidence for normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*, 161–177.

Kubovy, M., & Van Vanlkenburg, D. (2001). Auditory and visual objects. *Cognition*, *80*, 97–126.

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*, 143–151.

Moore, C., Yantis, S., & Vaughan, B. (1998). Object-based visual selection: Evidence from perceptual completion. *Psychological Science*, *9*, 104–110.

Neisser, U. (1967). *Cognitive Psychology*. New York: Appleton (Century/Crofts).

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

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltà, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favour of a premotor theory of attention. *Neuropsychologia*, *25*, 31–40.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*, 1–46.

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., Pavani, F., & Driver, J. (2000). Crossmodal links between vision and touch in covert endogenous spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1298–1319.

Stein, B. E., London, N., Wilkinson, L. K., & Price, D. P. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, *8*, 497–506.

Turatto, M., Benso, F., Galfano, G., & Umiltà, C. (2002). Nonspatial attentional shifts between audition and vision. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 628–639.

Turatto, M., Galfano, G., Bridgeman, B., & Umiltà, C. (2004). Space-independent modality-driven attentional capture in auditory, tactile and visual systems. *Experimental Brain Research*, *155*, 301–310.

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

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 crossmodal spatial attention orienting. *Perception & Psychophysics*, *62*, 1258–1264.

Wightman, F. L., & Jenison, R. (1995). Auditory spatial layout. In W. Epstein, & S. J. Roberts (Eds.), *Perception of space and motion*2nd ed. (pp. 365–400). San Diego, CA: Academic Press, 365–400.