When Cross-Modal Spatial Attention Fails

David J. Prime
Université de Montréal

John J. McDonald and Jessica Green
Simon Fraser University

Lawrence M. Ward
University of British Columbia

There is now convincing evidence that an involuntary shift of spatial attention to a stimulus in one modality can affect the processing of stimuli in other modalities, but inconsistent findings across different paradigms have led to controversy. Such inconsistencies have important implications for theories of cross-modal attention. The authors investigated why orienting attention to a visual event sometimes influences responses to subsequent sounds and why it sometimes fails to do so. They examined visual-cue-on-auditory-target effects in two paradigms—implicit spatial discrimination (ISD) and orthogonal cuing (OC)—that have yielded conflicting findings in the past. Consistent with previous research, visual cues facilitated responses to same-side auditory targets in the ISD paradigm but not in the OC paradigm. Furthermore, in the ISD paradigm, visual cues facilitated responses to auditory targets only when the targets were presented directly at the cued location, not when they appeared above or below the cued location. This pattern of results confirms recent claims that visual cues fail to influence responses to auditory targets in the OC paradigm because the targets fall outside the focus of attention.

Keywords: cross-modal attention, spatial attention, orthogonal cuing, implicit spacial discrimination, visual cuing of auditory attention

Numerous studies have demonstrated that spatially nonpredictive visual, auditory, or somatosensory cues can facilitate responses to nearby targets when the cues and targets are in different modalities (Kennett, Eimer, Spence, & Driver, 2001; McDonald, Teder-Sälejärvi, Di Russo, & Hillyard, 2003; McDonald, Teder-Sälejärvi, Heraldze, & Hillyard, 2001; McDonald, Teder-Sälejärvi, & Hillyard, 2000; McDonald & Ward, 2000; Schmitt, Postma, & De Haan, 2000, 2001; Spence & Driver, 1997; Ward, 1994; Ward, McDonald, & Lin, 2000). Such cross-modal cue effects support claims that involuntary shifts of spatial attention can be based on multimodal representations and that at least part of the neural system controlling those shifts may be supramodal. At one time, however, difficulties in observing cue effects between the auditory and visual modalities appeared to challenge the supramodal hypothesis (Driver & Spence, 1998). Although a variety of conflicting findings have been reported (e.g., Mondor & Amirault, 1998; Spence & Driver, 1997; Ward, 1994; Ward, McDonald, & Goleshani, 1998; Ward et al., 2000), there is growing evidence that spatially nonpredictive auditory cues do facilitate responses to nearby visual targets and vice versa (see Spence & McDonald, 2004, and Spence, McDonald, & Driver, 2004, for reviews). Nonetheless, a useful theoretical account of cross-modal spatial attention orienting awaits a more complete understanding of when cross-modal cuing succeeds and when it fails. In this article, we address this problem for two particular paradigms that have, in the past, yielded conflicting results.

These two paradigms, orthogonal cuing (OC) and implicit spatial discrimination (ISD), have been particularly useful for studying cross-modal cuing effects. In the OC paradigm, observers are required to discriminate the elevation of targets presented above or below fixation on either the left side or the right side of a display. Observers in the ISD task are required to respond to targets at some locations but to refrain from responding to targets at one or more other locations (McDonald & Ward, 1999). These paradigms share two features that are important for the study of audiovisual spatial attention. First, they require observers to engage in spatial processing. This feature appears to be critical for observing involuntary spatial attention-orienting effects on the processing of auditory targets, both for auditory cues and for cues from other spatial modalities. Auditory processing is initially nonspatial and is unaffected by spatial attention unless space is made relevant to the observer’s goals (McDonald & Ward, 1999; Rhodes, 1987; Spence, 2001). Second, unlike other tasks that make space relevant (e.g., left-right localization), both OC and ISD paradigms eliminate the potential problem of response priming by the cue. That is, a cue-induced bias toward making a response with the ipsilateral
hand cannot account for cuing effects in these paradigms, either because the same response is made on cued and uncued trials (ISD) or because the cued dimension and the target dimension on which the response is based are orthogonal (OC).

In the original ISD task, observers refrained from responding to targets presented at fixation and responded to targets presented elsewhere with a single button press (McDonald et al., 2001; McDonald & Ward, 1999, 2000; Ward et al., 2000). In other variants, the response to “go” targets also required a nonspatial discrimination (e.g., McDonald et al., 2001). The addition of the nonspatial discrimination allowed error rates to be examined to ensure that reaction time effects did not arise from differences in the response criterion between cued and uncued trials. Studies of this type have demonstrated that the appearance of a spatially nonpredictive visual cue facilitates responses to nearby auditory targets, and vice versa (McDonald, 1999). Moreover, event-related potential studies have demonstrated an enhancement of sensory-evoked neural activity for validly cued targets (McDonald et al., 2001; McDonald & Ward, 2000). These results provide convincing evidence that cross-modal attention effects do occur between auditory and visual cues and targets.

Originally used to study attention shifting within the auditory modality (Spence & Driver, 1994), the OC paradigm has also been used to demonstrate that the appearance of a spatially nonpredictive auditory cue facilitates responses to visual targets on the same side of the vertical meridian (Schmitt et al., 2000; Spence & Driver, 1997; Spence, Nicholls, Gillespie, & Driver, 1998). Several OC studies, however, have found that spatially nonpredictive visual cues fail to facilitate responses to auditory targets on the same side of the vertical meridian (Rorden & Driver, 1999; Schmitt et al., 2000; Spence & Driver, 1997). These results were originally taken as evidence for a “missing link” between separate but sometimes interactive visual and auditory attentional control systems (e.g., Driver & Spence, 1998).

In spite of considerable evidence for more-or-less-symmetric involuntary spatial attention effects between the auditory and visual modalities, it is still unclear exactly which factors determine whether such effects will appear in a particular paradigm. The aim of the present study, therefore, was to test a specific hypothesis about why involuntary attention to a visual cue sometimes facilitates responses to nearby auditory targets and sometimes fails to do so. Ward et al. (2000) proposed that the spatial separation between visual cue and auditory target must be relatively small for such effects to appear (see also Spence, 2001). The null effects found in the OC paradigm could have resulted from its requirement that the vertical distance between cues and targets be large (14° in previous experiments) so that observers could reliably discriminate target elevation. A visual cue might focus attention too narrowly at its location to affect the processing of auditory targets that appear far away from it. Support for this proposition comes from studies that have demonstrated spatial attention gradients within and between the visual and auditory modalities (Handy, Kingstone, & Mangun, 1996; Mondor & Zatorre, 1995; Schmitt et al., 2001).

We tested this narrow-focus hypothesis by comparing performance in an OC paradigm and an ISD paradigm with identical cues and targets. In the OC paradigm, auditory targets were presented from speakers positioned 15° above and below the visual displays that were used to present the cues. In the ISD paradigm, auditory targets were also presented from speakers positioned directly at the visual cue locations. If the narrow-focus hypothesis were correct, the visual cue would facilitate responses to ipsilateral auditory targets only in the implicit spatial discrimination paradigm and only when the target appeared directly at a cued location.

Method

Observers

All observers reported normal hearing and normal or corrected-to-normal vision. Fifteen observers performed in the OC paradigm and 22 performed in the ISD paradigm. One observer from the OC group was eliminated due to near-chance performance. Six observers were excluded from the ISD group, 3 because of a failure to maintain eye fixation and 3 because of high no-go false alarm rates (>20%). Fourteen observers (7 women and 7 men) remained in the OC group and 16 (9 women and 7 men) remained in the ISD group.

Stimuli and Procedure

Figure 1 shows the stimulus array used in both conditions. A start-of-trial warning signal consisted of a 50-ms burst of white noise (74 dBA; 5-ms rise–fall) from the center speaker and a simultaneous 50-ms flash of the eight nonfixation LEDs on the center speaker. Visual cues consisted of a 50-ms flash of all eight LEDs on either the left or the right side. An auditory target consisted of four 20-ms broadband noise bursts (2.5-ms rise–fall) separated by 10-ms silent intervals (total duration 110 ms). Target
stimuli were presented at either a “quiet” 66-dB (A) or “loud” 72-dB (A) intensity level.

Observers responded to targets by pressing the upper and lower response buttons with their right index and middle fingers. Response accuracy feedback was presented at the end of each block on a computer monitor positioned to the observer’s left.

The horizontal and vertical electro-oculogram was recorded, and semiautomated artifact rejection routines were used to discard trials during which blinks or eye movements greater than about 3° occurred in the interval from 100 ms before cue onset until 600 ms after target onset. After artifact rejection, average eye position, occurred in the interval from 100 ms before cue onset until 600 ms after target onset. After artifact rejection, average eye position, calculated separately for left and right visual field stimuli, deviated from fixation by less than 2 μV (about 0.2°) for all retained observers.

Observers were instructed to fixate the central LED and to blink, when necessary, between trials. In the OC task, each trial began with the presentation of the start-of-trial signal. After a 750-ms delay, the visual cue was presented with equal probability from the left or right side. Following a randomly variable stimulus onset asynchrony (SOA) of 100 ms or 500 ms, the auditory target was then presented with equal probability from one of the four possible target locations (upper or lower left or right). On half of the trials, the cue and target were presented on the same side of fixation. On the other half of the trials the cue and target were presented on opposite sides. Observers were instructed to press the upper button when a target was presented at one of the upper positions and to press the lower button for a quiet target. They completed seven blocks of trials (96 go trials and 24 no-go trials per block) and received both accuracy and no-go false alarm rate feedback between blocks.

Observers were instructed to press the upper button when the target was presented from the center speaker. Observers responded to targets from any peripheral speaker (go trials) and withheld responses from targets presented from the center speaker (no-go trials). Observers were told to press the upper button for a loud target and the lower button for a quiet target. They completed seven blocks of trials (96 go trials and 24 no-go trials per block) and received both accuracy and no-go false alarm rate feedback between blocks.

### Results

#### Orthogonal Cuing Paradigm

Trials with eye movements were discarded (14%). Reaction times (RTs) less than 100 ms or greater than 1,000 ms (about 3 standard deviations from the median) were also discarded (5.1% of trials). Incorrect responses were excluded from the RT analysis. Median RTs and error rates were calculated from the remaining data for each observer for the four SOA (100 ms or 500 ms) × cuing (cued-side or uncued-side) conditions. Table 1 shows overall mean RTs and error rates, and the cue effects (uncued side minus cued side) are depicted in Figure 2.

Separate analyses of variance (ANOVA) as repeated measures were performed on RTs and error rates. As expected, visual cues failed to facilitate performance on cued-side trials relative to uncued-side trials. The RT analysis failed to reveal any significant main effects or interactions (all ps > .05), although the main effect of cuing, $F(1, 13) = 4.10, p = .064$, and the SOA × Cuing interaction, $F(1, 13) = 4.10, p = .073$, did approach significance. Subsequent paired comparisons revealed that RTs were 15 ms shorter on uncued-side trials relative to cued-side trials at the 100-ms SOA, $t(13) = 2.83, p = .014$, the opposite of an attentional facilitation effect. No difference between cued-side and uncued-side trials was found at the 500-ms SOA, $t(13) = 0.18, p = .858$.

The analysis of error rates failed to reveal any significant main effects or interactions (all ps > .05). Consistent with the RT results, however, observers committed more errors on cued-side
(5.2%) than on uncued-side (3.2%) trials at the 100-ms SOA. Although this difference failed to reach statistical significance, $t(13) = 1.76, p = .10$, the direction of the effect indicates that the RT results did not arise from speed–accuracy tradeoffs.

**Implicit Spatial Discrimination Paradigm**

Observers responded to 6.1% of the no-go trials. Approximately 15% of go trials were discarded because of eye movements. RTs less than 100 ms or greater than 1,200 ms (about 3 standard deviations from the median) were discarded (3.8%), and incorrect responses were excluded from the RT analysis. Median RTs and error rates were calculated for each observer from the remaining data for the eight SOA (100 ms or 500 ms) × cuing (cued-side or uncued-side) × target elevation (middle, upper, and lower) conditions (see Table 1). As expected, RTs were significantly faster on cued-side than on uncued-side trials, $F(1, 15) = 6.96, p = .019$. Critically, there was also a significant interaction between cuing and target elevation, $F(1, 15) = 5.32, p = .036$. No other effects approached significance (all $ps > .10$). To examine the Cuing × Target Elevation interaction, separate 2 (cuing) × 2 (SOA) ANOVAs were calculated for targets presented at either the middle or upper and lower locations. Consistent with the narrow-focus hypothesis, target responses were facilitated when targets were presented at the cued middle speaker, $F(1, 15) = 6.85, p = .019$, but not when targets were presented 15° above or below the cued location, $F(1, 15) < 1.00$. No other effects approached significance (all $ps > .10$). The error rate analysis revealed a significant main effect of target elevation, $F(1, 15) = 11.36, p < .01$. More errors were committed for targets at the middle elevation. By contrast, there was essentially no difference in error rates between cued-side (14.8%) and uncued-side (14.7%) trials for targets presented at the upper and lower elevations. This indicates that the RT results did not arise from speed–accuracy tradeoffs.

**Discussion**

We investigated why orienting attention to a spatially nonpredictive visual cue sometimes has no influence on responses to subsequent auditory targets. Following Ward et al. (2000), we hypothesized that attention may be too narrowly focused around the location of a lateralized visual event to facilitate responses to auditory events that occur relatively far away. This narrow-focus hypothesis predicts that a lateralized visual cue would facilitate responses to auditory targets appearing at or near the cued location but not to targets appearing above or below the cued location.

The present results are consistent with the narrow-focus hypothesis. As predicted, we found that a spatially nonpredictive visual cue did facilitate responses to targets appearing at the cued location in the ISD paradigm, thereby replicating the results of previous experiments (McDonald, 1999; McDonald et al., 2001; Ward et al., 2000). Furthermore, this facilitatory cuing effect was obtained only for targets presented at the cued location. No differences in performance were observed between cued-side and uncued-side trials when targets were presented 15° above or below the cued location. This latter finding is consistent with the typical result obtained in OC experiments that present auditory targets at similar distances from visual cues (Rorden & Driver, 1999; Schmitt et al., 2000; Spence & Driver, 1997). The present results are also consistent with prior research that has examined the spatial gradient of cross-modal cue effects in the azimuthal plane. Schmitt et al. (2001) found that visual cues facilitated localization responses for auditory targets that appeared at the cued location but not at adjacent locations.
Interestingly, in the OC paradigm RTs were 15 ms longer on cued-side trials than on uncued-side trials when the SOA was 100 ms. Recently, Mazza, Turatto, Rossi, and Umiltà (2007) also found an inhibitory cue effect for auditory targets preceded by visual cues in an OC experiment. Furthermore, Ward et al. (1998, Experiment 8) found a similar inhibitory effect between auditory cues and visual targets in the OC paradigm when the SOA was shorter than 200 ms. The cause for the present visual-on-auditory inhibitory effect is unknown because there are several minor methodological differences between the present OC experiment and previous ones that did not find this effect. Nevertheless, it is clear that the appearance of a spatially nonpredictive visual cue does not facilitate processing of sounds appearing 15° above or below the cued location in either the OC paradigm or the ISD paradigm.

Because of the different ways in which visual and auditory systems process spatial information, the same task demands in the OC paradigm led to quite different cross-modal cue effects depending on which modality was the cue and which was the target. Conversely, a different paradigm, in this case ISD, revealed cross-modal cue effects regardless of cue and target modality. The ISD and OC tasks have been used extensively to study effects of spatial attention both within and between sensory modalities. Both tasks ensure that the target location is relevant, which is essential for observing spatial cue effects in audition while at the same time avoiding confounding effects of S-R compatibility and other methodological problems (see Spence & McDonald, 2004). The ISD task has the additional advantage that cue and target stimuli can be presented from exactly the same locations on cued-side trials. In contrast, the OC task normally involves a spatial separation between cues and targets. In the case of auditory targets, the spatial separation needed to perform the task is large because of the poor spatial acuity of the auditory system. The present findings indicate that in the case of visual cues and auditory targets, the difference in the spatial relationship between cue and target in the two paradigms is the key factor giving rise to the observed differences in results.

Résumé

Il y a aujourd’hui des preuves convaincantes qu’un déplacement involontaire de l’attention dans une modalité peut influencer le traitement des stimuli dans une autre modalité, mais des données contradictoires obtenues à l’aide de différents paradigmes sèment la controverse. De telles contradictions ont des implications importantes pour les théories de l’attention intermodale. Ici, nous avons examiné pourquoi porter attention à un événement visuel influence parfois la réponse à un bruit subséquent et d’autres fois non. Nous avons examiné les effets d’un indice-visual-sur-cible-auditif en fonction de deux paradigmes - la discrimination spatiale implicite (DSI) et l’indicateur orthogonal (IO) - qui ont mené à des résultats contradictoires dans le passé. Conformément aux recherches antérieures, les indices visuels ont facilité la réponse aux cibles auditives du même côté dans le paradigme de DSI mais pas dans le paradigme d’IO. En outre, dans le paradigme de DSI, les indices visuels ont facilité la réponse aux cibles auditives seulement quand les cibles étaient présentées directement à l’endroit indiqué, et non quand elles apparaissaient au-dessus ou au-dessous de l’endroit indiqué. Ce patron de résultats confirme les propositions récentes selon lesquelles les indices visuels n’influencent pas la réponse aux cibles auditives dans le paradigme d’IO parce que les cibles se trouvent en dehors du focus attentionnel.

Mots-clés : attention intermodale, attention spatiale, indigage orthogonal, discrimination spatiale implicite, indicage visuel de l’attention auditive

References


Received May 17, 2007
Accepted August 21, 2007