Exp Brain Res (2004) 155: 301–310
DOI 10.1007/s00221-003-1724-x

RESEARCH ARTICLES

Massimo Turatto · Giovanni Galfano · Bruce Bridgeman · Carlo Umiltà

Space-independent modality-driven attentional capture in auditory, tactile and visual systems

Received: 12 November 2002 / Accepted: 11 September 2003 / Published online: 5 December 2003

Abstract Extending previous evidence for attentional shifts across auditory and visual modalities without the confound of the two modalities originating at different locations (Turatto et al. 2002), we investigated attention shifts between auditory and tactile modalities, and between tactile and visual modalities. Two stimuli (S_1 and S_2), either in the same or in different modalities, were delivered from the same spatial source and were separated by a variable temporal gap. S_1 was task irrelevant, whereas S_2 required a speeded discrimination. Results showed that modality switching is detrimental independently of the stimulated modality as long as the temporal lag between S_1 and S_2 is short enough that there is not time to switch attention before S_2 is delivered. We observed automatic, modality-driven, attentional capture, with ipsimodal trials leading to faster response times than crossmodal trials. The present results cannot be accounted for by spatial artifacts, response priming or criterion shifts, and are interpreted as the consequence of a space-independent attentional shift across sensory modalities.

The research was supported by grants from CNR and MURST to C. U.

M. Turatto · G. Galfano
University of Trento,
Rovereto, Italy

B. Bridgeman
University of California,
Santa Cruz, USA

C. Umiltà
University of Padova,
Padova, Italy

M. Turatto (✉)
Department of Cognitive Science and Education, University of Trento,
Via Matteo del Ben 5,
38068 Rovereto, Italy
e-mail: massimo.turatto@unitn.it
Tel.: +39-0464-483558
Fax: +39-0464-483554

Keywords Space-independent attentional capture · Modality-driven attentional capture · Auditory systems · Tactile systems · Visual systems

Introduction

Research on crossmodal attention has recently aimed at understanding the mechanism by which attention integrates and coordinates information coming from different sensory modalities (e.g., Driver and Spence 1994, 1998; Stein et al. 1996). Such integration is necessary in that, apart from the experimental conditions devised in the laboratory to investigate stimulus processing in one modality at a time, our everyday life experience is multimodal rather than unimodal. Recent studies point to the existence of numerous crossmodal links in spatial attention (e.g., Driver and Spence 2000; Spence 2001; Stein et al. 1995).

Traditionally, the issue of spatial attention has been relegated to vision, with the majority of studies showing that a transient visual event like the onset of a brief flash of light (the cue) can produce an orienting of attention to its location. This is reflected by faster response time (RT) to a following similar visual event (the target) when it occurs at the same spatial location (or nearby) compared to when it occurs elsewhere in the display (Posner 1980). Mechanisms governing the allocation of attention in the visual field have been proved to be either automatic (e.g., Jonides 1981; Müller and Rabbitt 1989), or voluntary (e.g., Posner 1980; Posner et al. 1980), with the possibility of localizing the corresponding neural substrates in the parietal areas and the frontal areas respectively (Posner and Petersen 1990).

Automatic and voluntary orienting of spatial attention has also been observed when the cue and the target are delivered from different modalities (e.g., Caclin et al. 2002; Eimer and Schröger 1998; Schmitt et al. 2000; Spence and Driver 1996, 1997a, 2000; Spence et al. 2000a, 2000b). By means of the ‘orthogonal-cueing’ paradigm, Spence and Driver (1997a) provided evidence that, on the whole, a spatial cue in a given modality (e.g., audition) can capture attention to its location either in the same or in a different modality (e.g., vision).

Besides these recent studies emphasizing the role of attention across modalities in the spatial domain, the relationship between sensory modalities and attention was also investigated, especially in the 70 s and 80 s, from a different perspective. At issue were the non-spatial shifts of attention across modalities, and in particular the attentional mechanisms for stimulus selection when two stimuli (S_1 and S_2) are delivered from the same or different modalities, separated by a brief temporal interval (e.g., Harvey 1980; Klein 1977; Posner et al. 1976). Overall, the results showed RTs on crossmodal trials to be slower than RTs on ipsimodal trials, because attention had to be switched across modalities in the former but not in the latter condition. An exception to this rule was detected when observers were to respond to a visual S_2 after an auditory S_1. In this case, RTs were faster on crossmodal trials (auditory S_1 followed by a visual S_2) than on ipsimodal trials (visual S_1 followed by a visual S_2), an effect that could be due to the automatic alerting effect of auditory stimuli (e.g., Harvey 1980; Posner et al. 1976) and the generally faster response to auditory than to visual transients.

However, as pointed out by Spence and Driver (1997b), these early studies presented a series of methodological flaws such as response priming, criterion shifts, and the spatial artifact, that render it impossible to interpret the results as pure non-spatial attentional effects. In particular, the spatial artifact, which was common in many early studies, was due to the fact that stimuli from different modalities were delivered from different spatial sources. For example, whereas the visual stimuli were administered through a CRT display located in front of the participant's head, auditory stimuli were delivered through headphones, leading to a spatial mismatch between the two sources of stimulation. Hence, on crossmodal trials, two kinds of attentional shifts presumably occurred: One from the modality of S_1 to the modality of S_2, and one from the location of S_1 to the location of S_2. It follows that the slower RTs observed on crossmodal trials as compared to ipsimodal trials might result from confounded effects of a shift of attention across modalities and of a spatial shift of attention between two different sources of stimulation.

In a recent study, avoiding the methodological artifacts previously identified by Spence and Driver (1997b), Turatto et al. (2002) re-examined the issue of non-spatial shifts of attention across auditory and visual modalities. In particular, they hypothesized that when a stimulus is perceived, attention is briefly and automatically captured to the corresponding modality, thus favoring the processing of a subsequent ipsimodal stimulus compared to a crossmodal stimulus. Evidence for a non-spatial attentional modulation across modalities was also provided in a positron emission tomography (PET) study by Kawashima et al. (1995), who measured regional cerebral blood flow (rCBF) from visual cortex while participants were engaged in somatosensory tasks. On the assumption that attentional modulation of cerebral metabolic rate is reflected in changes of the rCBF (e.g., Corbetta et al. 1991), the results indicated that attentional deployment to tactile modality caused a reduction of the rCBF in the visual cortical areas, which occurred regardless of whether participants performed the somatosensory task with their eyes open or closed. Thus, in addition to the well documented crossmodal links in spatial attention (see Spence 2001 for a recent review), the issue of a non-spatial attentional modulation across modalities also seems worthy of further investigation, especially considering that the studies that have been conducted so far are very often susceptible to methodological criticisms (see Spence and Driver 1997b).

Turatto et al. (2002) used the intersensory facilitation paradigm (Harvey 1980), consisting of the presentation of two consecutive stimuli S_1 and S_2, separated by a variable temporal gap, both in the same modality and in different modalities. Specifically, the authors were interested in audition and vision. Auditory and visual stimuli were administered through the same device, so that both kinds of stimuli originated from the same spatial coordinates, thus avoiding the spatial artifact. Participants were required to perform either a simple detection task (Experiments 1–4) or a choice RT task (Experiments 5 and 6) on S_2. In the former case, they were to respond as quickly as possible to the onset of the target regardless of its modality. The target, which could be either a light-emitting diode (LED) flashed in green or a 1,800-Hz pure tone, followed the onset of the cue after a variable delay (150, 600 or 1,000 ms). Overall, the results indicated that participants reacted faster when S_1 and S_2 were in the same modality compared to when they were in different modalities. When cue modality was informative of the target modality the advantage of ipsimodal trials over crossmodal trials could be explained assuming that, upon S_1 presentation, participants endogenously directed their attention to the cued modality.

Crucially, such an ipsimodal advantage was found even when S_1 and S_2were uncorrelated (Experiment 1), which, in our view, points to the fact that attention was exogenously grabbed by S_1 modality. If the upcoming S_2 was presented in the same modality, then RTs were faster compared to when $S_2$was delivered in a different modality, because attention in the latter condition had to be shifted across modalities (Turatto et al. 2002). The automatic nature of such capture is further testified by the fact that the ipsimodal advantage was present only at the shortest (150 ms) stimulus onset asynchrony (SOA), fading away at the longer SOAs (600 and 1,000 ms). This short-latency effect is typical of many other automatic processes, for example, the exogenous orienting of spatial attention (e.g., Jonides 1981; Müller and Rabbitt 1989) and Stroop-like interference (e.g., Logan 1980).

Although the analysis of the false-alarm rate (i.e., response on catch trials) in the simple RT task adopted by Turatto et al. (2002, Experiments 1 and 4) rendered it implausible that a criterion shift affected the results, nonetheless one might argue that the use of a choice RT task is a safer way to rule out the possibility that data were contaminated by this non-attentional phenomenon (Spence and Driver 1997b). Hence, to provide stronger evidence on the exogenous-attentional nature of the effect, Turatto et al. replicated the same experiments with participants performing a choice RT task. When presented with a visual S_2, participants had to discriminate between the LED flashing in red vs. green, whereas when presented with an auditory S_2, participants were to discriminate between a high-pitch tone vs. a low-pitch tone. The findings confirmed the stimulus-driven attentional capture observed with the simple RT task. In addition, Turatto et al. provided evidence for such capture to be mandatory, in that the ipsimodal advantage emerged even when S_2 modality was blocked, i.e. when participants knew with 100% certainty the modality of the upcoming target, therefore having the possibility to fully focus their attention on its modality in advance.

In the present study, space-independent attentional shifts across modalities were extended to touch. Specifically, auditory vs. tactile and tactile vs. visual conditions were compared. We used the rationale and procedure adopted in Experiments 5 and 6 of Turatto et al.'s (2002) study, requiring a choice RT. Hence, attentional capture across modalities was evaluated either when participants had no information about the target modality (S_2 uncertain, Experiments 1 and 3), or when participants knew the target modality in advance (S_2 100% certain, Experiments 2 and 4). In the former condition, capture was tested when observers had no incentive to pay attention to a particular modality, whereas in the latter condition we tested the mandatory nature of such capture, in that participants had the opportunity to prevent involuntary ipsimodal capture, if any, by fully focusing attention on the target modality prior to target occurrence. In both cases, S_1 modality was uncorrelated with S_2 modality. In Experiments 1 and 2 we compared audition and touch. Experiments 3 and 4 were concerned with vision and touch. We want to make clear that in the present set of experiments although attention was spatially allocated, results cannot be accounted for by invoking shifts of spatial attention because stimuli from the different modalities were provided by the same object.

Experiment 1

This experiment was aimed at investigating whether attention is automatically captured by the modality of S_1 when participants are requested to process a subsequent S_2, and both stimuli are delivered to the same location. If this were the case, we expected S_1 to produce RT advantages in processing a subsequent ipsimodal S_2 compared to a crossmodal S_2. Because S_1 modality was not predictive of S_2 modality, participants could not use any deliberate strategy to allocate attention to S_1modality, and thus the role of expectancy-endogenous factors should be minimized, if not totally excluded. The same SOAs between S_1 and S_2 as those used by Turatto et al. (2002) were used.

Materials and methods

Participants

Eighteen psychology students from the University of Padova served as participants (12 females and 6 males). All were unaware of the purpose of the experiment and gave their informed consent prior to their participation. Their mean age was 23.1 years (range 20–31 years). All were right-handed by self-report and reported normal hearing and normal or corrected-to-normal vision.

Apparatus and stimuli

All experiments were conducted in a darkened room with a background luminance of 0.5 cd/m^2. The subject sat in front of a table facing the modality-stimulator device, composed of a circular loudspeaker cone (3 cm, 8 Ω, 200 mW, RS 248–476) and an LED (1.15° of visual angle). The loudspeaker was mounted just behind the LED in a plastic box. Just below the LED (1.5 cm) there were two openings in the box, where participants introduced their index and middle finger-pads of the left hand. The apparatus for the generation of the tactile stimuli, which was also mounted inside the box, consisted of two miniaturized solenoids (3 W, ±12 V) with moving cylindrical metallic plungers (1.4 mm in diameter, 50 mm in length), oriented perpendicularly to the pad of each finger. When activated, the plunger of a given stimulator moved 4 mm vertically for 100 ms, touching the finger-pad.

For the auditory modality, S_1 was a 900+1,800-Hz pure tone presented at 70 dB(A), whereas S_2 was either a 900-Hz or a 1,800-Hz pure tone presented at 70 dB(A).

For the tactile modality, S_1 was a touch of the middle finger of the left hand, whereas S_2 was either a touch of the index or the middle finger pads of the left hand.

Design

Participants were tested in a 2×2×3 factorial design. The first factor was S_1 modality (auditory vs. tactile). The second factor was S_2 modality (auditory vs. tactile). We will refer to the four conditions as: AA (auditory S_1 followed by auditory S_2), AT (auditory S_1 followed by tactile S_2), TT (tactile S_1 followed by tactile S_2), and TA (tactile S_1 followed by auditory S_2). The third factor was SOA (150, 600, or 1,000 ms).

The total number of trials was 192, divided into 4 experimental blocks of 48 trials each. Specifically, there were 16 trials for each SOA, 8 trials for the ipsimodal S_1-S_2 combination, and 8 for the crossmodal S_1-S_2 combination.

Procedure

Each participant was fully informed that S_1 modality was uncorrelated with S_2 modality. Practice trials were provided until the participant felt confident with the task. On each trial, S_1 and S_2 were delivered for 100 ms, separated by the appropriate SOA. Participants had to react as quickly as possible by using the right hand and pressing the 'H' key of the keyboard with their middle finger when S_2 was the touch of the middle-finger pad or the low-pitch tone, and the 'B' key of the keyboard with their index finger when S_2 was the touch of the index-finger pad or the high-pitch tone.

Table 1 Mean error rates for the different S1-S2 combinations at each SOA in Experiments 1, 2, 3, and 4

	SOA 150				SOA 600				SOA 1,000			
	AA	AT	TT	TA	AA	AT	TT	TA	AA	AT	TT	TA
Exp. 1	6.6	4.9	5.4	5.1	6.2	6.2	8.8	4.8	6.1	5.3	6.5	7.8
Exp. 2	5.5	3.8	6.4	6.9	5.0	5.1	3.0	4.1	4.1	6.3	3.8	4.3
	TT	VT	VV	TV	TT	VT	VV	TV	TT	VT	VV	TV
Exp. 3	4.5	4.3	4.2	6.3	6.6	7.1	2.4	3.9	4.0	3.9	6.0	6.2
Exp. 4	7.8	3.5	4.2	6.0	8.9	4.1	4.5	7.1	3.6	4.4	5.0	3.6

Results

Error analysis

Overall error rate was less than 7% (see Table 1). An ANOVA with S_1 modality, S_2 modality, and SOA as factors did not reveal any statistical differences in error distribution across conditions. In particular, the second order interaction was not significant (F<1).

RT analysis

In this and in the following experiments, before analysis was carried out, RTs for correct responses were trimmed according to the method developed by Van Selst and Jolicoeur (1994). As a consequence of the outlier-latency criterion, 1.7% of the data were trimmed. RT data were entered into a repeated measures ANOVA in which the factors were those already mentioned. The main effects of S_1 modality $F_{(1,17)} = 5.474$, p <.04, S_2 modality $F_{(1,17)} = 14.004$, p <.002, and SOA $F_{(2,34)} = 24.236$, p <.001, were significant.

Both the S_1 modality × S_2 modality interaction, $F_{(1,17)} = 14.363$, p <.001, and the S_1 modality × S_2 modality × SOA interaction, $F_{(2,34)} = 6.273$, p <.001, were significant. In order to qualify the second order interaction, planned comparisons were applied to the data. As shown in Fig. 1, at the shortest SOA (150 ms), for both modalities, RTs were faster on ipsimodal trials than on crossmodal trials (AA: M = 684 ms, SD = 147, TA: M = 751 ms, SD = 174, p <.04; TT: M = 766 ms, SD = 153, AT: M = 868 ms, SD = 200 p <.001). At the intermediate SOA (600 ms), ipsimodal trials were faster than crossmodal trials in the tactile modality only (TT: M = 688 ms, SD = 131, AT: M = 754 ms, SD = 143, p <.001). No significant differences between ipsimodal and crossmodal trials emerged at the longest SOA (1,000 ms).

Discussion

The RT pattern which emerged in the present experiment is fully consistent with our main prediction. In fact, we hypothesized that, in a S_1-S_2 paradigm, as soon as S_1 is presented, attention automatically and briefly shifts to the stimulated modality. These results confirm and extend previous findings showing the same effect across visual and auditory modalities (Turatto et al. 2002).

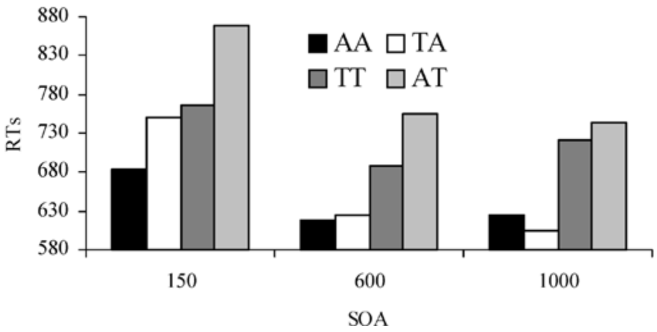

Fig. 1 The S_1 modality × S_2 modality × SOA interaction in Experiment 1. S_1 modality was not informative about S_2 modality (*TT* tactile S_1 followed by a tactile S_2, *TA* tactile S_1 followed by an auditory S_2, *AA* auditory S_1 followed by an auditory S_2, *AT* auditory S_1 followed by a tactile S_2)

Two aspects point to the exogenous nature of the phenomenon observed. First, because S_1 modality was uninformative of S_2 modality, participants had no incentive to voluntarily attend to the modality of the first stimulus. Second, similarly to what happens in the orienting of spatial attention (e.g., Jonides 1981; Müller and Rabbitt 1989), the advantage observed on valid trials (i.e., ipsimodal trials in this experiment) over invalid trials (i.e., crossmodal trials) seems a short-latency phenomenon, being reliably present at the 150-ms SOA only. To summarize, the present findings show that whenever a task-irrelevant auditory or tactile stimulus (S_1) is presented, attention is automatically summoned to the corresponding modality. This space-independent, modality-driven, attentional capture turns out to be detrimental if the following target stimulus (S_2) is crossmodal compared to when it is ipsimodal, though this effect seems to be confined to the first 150 ms from stimulus presentation. An exception to this temporal constraint seems to emerge for the auditory modality, in that an auditory S_1 still affects the discrimination of a tactile target-S_2 600 ms later. We will discuss this finding in more detail in the "General discussion."

Experiment 2

As attested by the results of Experiment 1, upon S_1 presentation, attention is briefly grabbed by its modality, regardless of whether the observer has any intention of paying attention to it. However, as Turatto et al. (2002) pointed out, when S_2 modality cannot be predicted on the basis of S_1 modality, participants cannot focus their attention on the correct target modality prior to the occurrence of S_1, which might possibly prevent attention from being captured. To test the strength of such capture, however, one of the crucial conditions, beside load insensitivity and awareness, is that in which the criterion of *intentionality* could be, eventually, satisfied or instead violated (Jonides 1981). Hence, in Experiment 2, participants knew in advance, with 100% certainty, the modality of S_2 in a given block of trials, which gave them the possibility of fully focusing their attention on S_2 modality prior to the occurrence of S_1. If the modality of the first stimulus also affects RT for the target discrimination under such conditions, then one can reasonably conclude that the criterion of intentionality is met, and therefore the space-independent, modality-driven, capture of attention can be defined as automatic and mandatory.

Materials and methods

Participants

Eighteen psychology students of the University of Padova (11 females and 7 males) participated, and gave their informed consent prior to their participation. Their mean age was 24.6 years (range 22–26 years). All were right-handed by self-report and reported normal hearing and normal or corrected-to-normal vision. None of them had taken part in Experiment 1.

Apparatus and materials

The apparatus and materials were as in the previous experiment.

Design and procedure

The experiment proceeded as in Experiment 1, except that participants knew in advance, and with no uncertainty, the modality of S_2 in a given block of trials. S_1modality was congruent in half of the trials, and was incongruent in the remaining half. Order of presentation of S_2 modality was counterbalanced across participants.

The total number of trials was 192, divided into 4 experimental blocks of 48 trials each. Specifically, there were 16 trials for each SOA, 8 for the ipsimodal S_1-S_2 combination, and 8 for the crossmodal S_1-S_2 combination.

Results

Error analysis

Overall error rate was less than 5% (see Table 1). An ANOVA with the above factors did not reveal any statistical differences in error distribution across conditions, with the second order interaction remaining non-significant (F<1).

RT analysis

As a consequence of the outlier-latency criterion, 1.7% of the data were trimmed. RT data were entered into a repeated measures ANOVA in which the factors were the same as in Experiment 1. Neither the main effect of S_1 nor the main effect of S_2 modality were significant, though the main effect of SOA was significant, $F_{(2,34)} = 100.288$, $p < .001$.

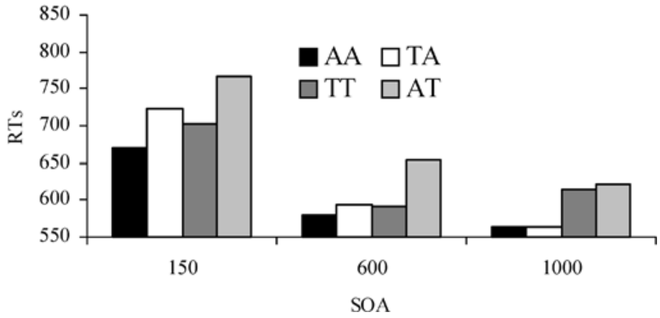

Fig. 2 The S_1 modality × S_2 modality × SOA interaction in Experiment 2. S_2 modality was always 100% valid, whereas S_1 modality was either congruent (50%) or incongruent (50%). The conventions are as in Fig. 1

Both the S_1 modality × S_2 modality interaction, $F_{(1,17)} = 5.534$, $p < .04$, and the S_1 modality × S_2 modality × SOA interaction, $F_{(2.34)} = 5.521$, $p < .009$, were significant. To qualify the second order interaction, planned comparisons were applied to the data. As shown in Fig. 2, at the 150-ms SOA, for both modalities, RTs were faster on ipsimodal trials than on crossmodal trials (AA: M = 671 ms, SD = 100, TA: M = 723 ms, SD = 134, $p < .003$; TT: M = 703 ms, SD = 143, AT: M = 768 ms, SD = 191, $p < .05$). As in Experiment 1, the 600-ms SOA still revealed a significant difference in RTs between ipsimodal and crossmodal trials for target-S_2 discrimination in the tactile modality (TT: M = 592 ms, SD = 92, AT: M = 653 ms, SD = 155, $p < .025$). At the 1,000-ms SOA ipsimodal trials and crossmodal trials were not significantly different.

Discussion

As Fig. 2 shows, the results closely paralleled Experiment 1 findings, showing an exogenous, modality-driven attentional capture at the 150-ms SOA. However, something more can be argued about the nature of such an involuntary shift of attention toward the modality of S_1. In fact, because participants had the opportunity to concentrate and allocate their attention to the ‘expected’ S_2modality, the very fact that despite this favorable condition their attentional resources were captured by S_1 modality is an index that this stimulus-driven mechanism appears to be strongly automatic and mandatory (see also Turatto et al. 2002). In other words, there is evidence that in the present S_1-S_2 paradigm, when the temporal gap between the stimuli is very brief, attentional allocation across modalities is heavily affected by exogenous factors such as S_1 modality rather than being under endogenous control, namely the observer’s expectation about S_2 modality.

Interestingly, results from Experiments 1 and 2 indicate that, in the auditory domain, this exogenous control of attentional allocation lasts more than 600 ms from S_1 presentation. Our data do not give a precise value for the duration of this exogenous allocation, only a range between 600 ms, where it is strongly present, and 1,000 ms, where it is not. We can note here that the longer auditory interference may be related to the temporal nature of the auditory modality, where meaning (for instance in language) is often carried in temporal structures that have durations of at least 600 ms. Our 600-ms minimum attentional duration corresponds to two or three syllables of normal speech, and speech pauses signaled by ‘uh’ and um’ in continuous English discourse are of about this duration (520 ms after ‘uh’ and 800 ms after ‘um’; Clark and Fox Tree 2002). The auditory stimulus has no instantaneous definition, being only a sound pressure level—it is defined only over time, unlike visual and tactile stimuli. However, as will emerge in Experiment 3, because a task-irrelevant visual S_1 also appears to be still detrimental 600 ms after its presentation for performing the same tactile discrimination, an alternative hypothesis about possible asymmetries between audition, vision and taction in capturing attentional resources is provided in the “General discussion.”

In the next two experiments, by using the rationale, design and procedure adopted in the previous experiments, we explored the space-independent attentional allocation across tactile and visual modalities.

Experiment 3

The logic behind the present experiment was identical to that of Experiment 1. Hence, attentional allocation across tactile and visual modalities was studied without the participants being informed about the modality of S_2. S_1 modality and S_2 modality were uncorrelated.

Materials and methods

Participants

Eighteen psychology students from the University of Padova served as participants (14 females and 4 males) and gave their informed consent prior to their participation. They were unaware of the purpose of the experiment and had not taken part in any of the previous experiments. Their mean age was 25.5 years (range 21–29 years). All were right-handed by self-report and reported normal hearing and normal or corrected-to-normal vision.

Apparatus and stimuli

For the tactile modality, the same stimuli as those in Experiments 1 and 2 were used. For the visual modality, S_1 was the onset of the LED in orange (CIE coordinates, x = .538, y = .416; 385 cd/m^2), whereas S_2 was the onset of the same LED in green (CIE coordinates, x = .424, y = .572; 232 cd/m^2) or red (CIE coordinates, x = .675, y = .324; 140 cd/m^2).

The total number of trials was 192, divided into 4 experimental blocks of 48 trials each. Specifically, there were 16 trials for each SOA, 8 trials for the ipsimodal S_1-S_2 combination, and 8 for the crossmodal S_1-S_2 combination.

Design and procedure

Design and procedure were the same as in Experiment 1 except that participants responded as quickly as possible by pressing the ‘H’ key when S_2 was the touch of the middle-finger pad or the LED was colored green, and the ‘B’ key when S_2 was the touch of the index-finger pad or the LED colored red.

Results

Error analysis

Overall error rate was less than 5% (see Table 1). An ANOVA with the above factors did not reveal any statistical differences in error distribution across conditions, the second order interaction being non-significant (F<1).

RT analysis

As a consequence of the outlier-latency criterion 1.6% of the data were trimmed. RT data were entered into a repeated measures ANOVA with the factors already mentioned. The main effects of S_1 modality, $F_{(1,17)}$ = 5.614, p <.03, S_2 modality, $F_{(1,17)}$ = 13.613, p <.002, and SOA, $F_{(2,34)}$ = 38.038, p <.001, were significant.

Both the S_1 modality × S_2 modality interaction, $F_{(1,17)}$ = 33.256, p <.001, and the S_1 modality × S_2 modality × SOA interaction, $F_{(2,34)}$ = 19.161, p <.001, were significant. As shown in Fig. 3, planned comparisons revealed that, at the shortest SOA (150 ms), for both modalities, RTs were faster on ipsimodal trials than on crossmodal trials (TT: M = 624 ms, SD = 111, VT: M = 778 ms, SD = 138, p <.04; VV: M = 701 ms, SD = 92, TV: M = 770 ms, SD = 106, p <.001). At the intermediate SOA (600 ms), ipsimodal trials were faster than crossmodal trials only in the tactile modality (TT: M = 561 ms, SD = 82, VT: M = 602 ms, SD = 100, p <.005) and again the magnitude of the difference was reduced compared to the shorter SOA condition. No significant differences between ipsimodal and crossmodal trials emerged at the 1,000-ms SOA.

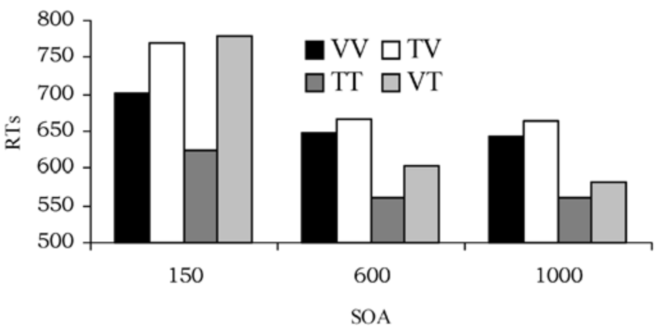

Fig. 3 The S_1 modality × S_2 modality × SOA interaction in Experiment 3. S_1 modality was not informative about S_2 modality (*TT* tactile S_1 followed by a tactile S_2, *TV* tactile S_1 followed by a visual S_2, *VV* visual S_1 followed by a visual S_2, *VT* visual S_1 followed by a tactile S_2)

Discussion

Basically, the results paralleled those of Experiment 1, showing that even though the modality of the first stimulus was uncorrelated with that of the target, when the temporal lag between S_1 and S_2 was very brief, RTs for S_2 discrimination were affected by S_1 modality. In particular, it appears that either a visual or a tactile stimulus, once presented, automatically engages attention to its modality, thus facilitating the discrimination of a following ipsimodal target compared to a crossmodal target. This result, however, appears to be short lived, in that it seems to vanish when the temporal lag between S_1 and S_2 is longer than 150 ms (also see Turatto et al. 2002). As for the auditory modality in both Experiments 1 and 2, an exception to this pattern emerges for the visual modality. In fact, the exogenous control of attention of a visual S_1 lasts up to 600 ms, causing slower response in the tactile discrimination task when compared to ipsimodal trials.

As we did for audition and touch in Experiment 2, in the next experiment we explored the degree of automaticity of the capture observed at the shortest SOA between tactile and visual modality by blocking the S_2 modality. We explore automaticity with respect to whether the ipsimodal advantage is influenced by knowing the modality of S_2 in advance.

Experiment 4

Materials and methods

Participants

Twenty psychology students of the University of Padova (14 females and 6 males) participated and gave their informed consent prior to their participation. Their mean age was 23.3 years (range 21–28 years). All were right-handed by self-report and reported normal hearing and normal or corrected-to-normal vision. None of them had taken part in the previous experiments.

Apparatus and materials

Apparatus and materials were the same as in Experiment 3.

Design and procedure

The design was the same as in Experiment 2. The procedure was as in Experiment 3 except that participants knew in advance, and with 100% certainty, the modality of S_2 in a given block of trials.

The total number of trials was 192, divided into 4 experimental blocks of 48 trials each. Specifically, there were 16 trials for each SOA, 8 trials for the ipsimodal S_1-S_2 combination, and 8 for the crossmodal S_1-S_2 combination.

Results

Error analysis

Overall error rate was less than 6% (see Table 1). An ANOVA did not reveal any statistical differences in error distribution across conditions, with the second order interaction being non-significant (F<1).

RT analysis

As a consequence of the outlier-latency criterion 1.9% of the data were trimmed. RT data were entered into a repeated measures ANOVA in which the factors were those reported in the previous experiments. The main effects of S_1 modality, $F_{(1,19)}$ = 7.656, p <.015, S_2 modality, $F_{(1,19)}$ = 14.846, p <.002, and SOA, $F_{(2,38)}$ = 63.426, p <.001, were significant. The S_1 modality × S_2 modality × SOA interaction, $F_{(2,38)}$ = 12.989, p <.001, was also significant. Planned comparisons applied to the data revealed an asymmetric attentional capture effect across visual and tactile modalities at the 150-ms SOA. In fact, as clearly emerges from the inspection of Fig. 4, RTs for the discrimination of S_2 in the visual modality were not affected by whether S_1 was itself a visual stimulus or a touch applied to the finger pad (VV: M = 576, SD = 97, TV: M = 580, SD = 122, n.s.). By contrast, RTs for discriminating which of the two finger pads was stimulated upon S_2-target appearance were clearly lengthened if the touch followed a visual S_1 rather than a tactile S_1 (TT: M = 602 ms, SD = 124, VT: M = 682 ms, SD = 115, p <.001). Moreover, a significant difference emerged showing that at the 600-ms SOA, for the visual modality, ipsimodal trials were slower than crossmodal trials (VV: M = 493 ms, SD = 71, TV: M = 469 ms, SD = 85, p <.03). This result is rather unexpected, and at the moment it is difficult to see how to interpret this finding. One trivial possibility is that it is due to random error. No other significant statistical differences emerged at either the 600- or the 1,000-ms SOA. Note that, as compared to Experiment 3, in the present experiment we failed to observe any effect of a visual S_1 over a tactile S_2 at the

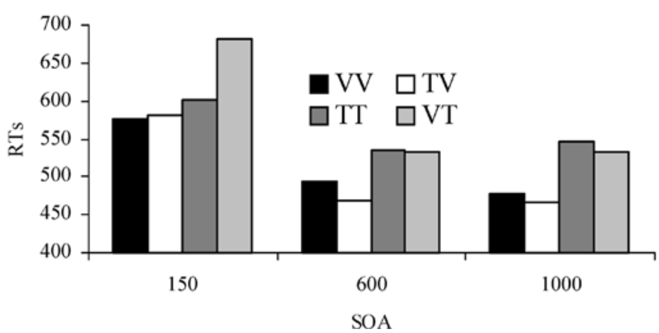

Fig. 4 The S_1 modality × S_2 modality × SOA interaction in Experiment 4. S_2 modality was always 100% valid, whereas S_1 modality was either congruent (50%) or incongruent (50%). The conventions are as in Fig. 3

600-ms SOA. We can only speculate that the 100% validity condition of the present experiment allowed the participant to exert better attentional control over incoming information, thus limiting the distracting effect of an irrelevant crossmodal visual S_1 over a tactile S_2 to the shortest SOA (150 ms).

General discussion

In recent papers on crossmodal links in spatial attention (see Spence 2001), it has been noted repeatedly that in the effort to understand the mechanisms by which selective attention operates, researchers in the last 40 years have concentrated their experiments on vision. However, because in everyday life we have to deal with a multisensory environment, the important issue of how attention picks up and coordinates information across different modalities (in particular audition, taction and vision) needs to be considered a central topic. With this purpose in mind, in the past decade, several research groups have begun to investigate this issue, giving rise to a great bulk of behavioral and electrophysiological evidence for the existence of several links in crossmodal spatial attention (Eimer and Schröger 1998; Driver and Spence 1998; Kennett et al. 2001; Stein et al. 1995; Tassinari and Berlucchi 1995; Tassinari and Campara 1996; Ward et al. 1998). In addition to some specific differences between exogenous and endogenous mechanisms of attention allocation, these studies extensively showed that when attention is directed toward a spatial location in a given modality, it becomes spatially oriented to the same (or a nearby) position in the other modalities also (e.g., Spence and Driver 1997a; but see Ward et al. 2000).

Despite the great effort made by these crossmodal spatial-attention studies in elucidating how attention integrates information from different sensory modalities, the potentially important issue of how attention selects information from different modalities coming from the same spatial location has been somehow disregarded in the literature (Spence 2002; Turatto et al. 2002). Early studies by Posner et al. (1976) and Klein (1977; also see Harvey 1980) investigated this issue more directly. Unfortunately, these pioneering studies presented a series of methodological flaws (i.e., response priming, criterion shift, spatial confound) that rendered the results equivocal as far as the role of attention was concerned.

By avoiding previous methodological confounds, in the present work we have sought to address non-spatial attentional shifts across audition and taction, and across taction and vision. Basically, the results expand upon the findings of our previous work where audition and vision were investigated (Turatto et al. 2002), by showing that when the temporal lag between S_1 and S_2 was very brief, a task-irrelevant S_1 grabbed attention to its modality, leading to a faster S_2 discrimination when S_1 and S_2 shared the same modality than when they did not. Because both stimuli were delivered from the same spatial source, the shift of attention should be considered space independent; instead, it would be purely modality driven. When such an attentional deployment turned out to be crossmodal with respect to the modality of the target, latency in target discrimination was greater than when attention was already allocated to the correct modality at the time the target was presented. As the studies of Kawashima et al. (1995) demonstrated, the allocation of attention to a given modality might result in both an increased neural activity to the attended modality-specific cortex and in a diminished activity in processing stimuli in the non-attended modality. One might wonder, however, whether the present results can be accounted for by invoking a task shift rather than by an attentional shift. According to the task-shift hypothesis, the advantage of ipsimodal trials over crossmodal trials would be due to the fact that the task (i.e., S_2 discrimination) differed in each modality, and, therefore, an S_1 in a given modality might have prepared participants to perform the task in the correspondent modality. On crossmodal trials a task shift occurred, whereas on ipsimodal trials it did not. This interpretation, however, is undermined by two observations. First, when the modality of S_2 was blocked, participants had no uncertainty about the type of task they were to perform, and, despite this, ipsimodal trials were faster than crossmodal trials at least at the shortest SOA. Second, if a task shift were involved, then it should be visible even at longer SOAs (Rogers and Monsell 1995). The possibility of a cost of switching, however, cannot be completely excluded in interpreting our data. Future research might address the switching cost issue directly.

Results from Experiments 1 and 3, in which the modality of the first stimulus was uncorrelated with that of the second (50% validity), revealed that when the lag time between the two stimuli was very brief (150-ms SOA), attentional deployment across audition and taction and across taction and vision was under exogenous control. When the strength of such automatic capture was assessed, Experiment 2 showed that it was mandatory with respect to the relationship between audition and taction. Therefore, even when participants knew in advance that the S_2 modality remained fixed in a given block of trials, a crossmodal S_1 delayed RTs for S_2 discrimination compared to an ipsimodal S_1. By contrast, participants were not distracted by an irrelevant tactile S_1 when fully attending to vision, whereas their RTs for a tactile S_2 discrimination slowed down when a tactile target briefly followed a visual S_1, as compared to when it followed a tactile S_1 (see Experiment 4). Another interesting aspect emerging from Experiments 1 and 3 was that the exogenous control of attention seemed to last longer for both audition and vision compared to touch. In fact, at the 150-ms SOA attentional capture was symmetrical between audition and touch, and between touch and vision. At the 600-ms SOA, there was still a crossmodal effect for the tactile S_2.

Comparing the experiments with an unpredictable response modality (Experiments 1 and 3) to the experiments with a predictable modality (Experiments 2 and 4), we can see that for the eight modality combinations and the three SOAs the responses to a predictable modality were faster than the responses to the corresponding unpredictable modality in every case. The magnitude of the advantage, however, was uneven. In most cases the lower ipsimodal RT was seen in both predictable and unpredictable conditions. The results of Experiments 3 and 4 show that predictability of the modality shortened the ipsimodal visual RT, but shortened the crossmodal visual RT even more. This effect was so strong that it even reversed the ipsimodal advantage for visual over tactile at the 600-ms SOA (Fig. 4), though by a quantitatively small amount. Apparently, when the participant knows what the response will be to the visual modality, the modality of S_1 has only small and inconsistent effects.

Different reasons could perhaps explain these asymmetries between audition and touch and between touch and vision in capturing attention and in keeping it anchored to the correspondent modality. First, it should be noted that tactile stimuli are necessarily registered in the personal space, whereas the auditory and visual systems are particularly equipped to encode stimuli coming from the peripersonal, or, even more frequently, from the extrapersonal space. In human beings, as well as in other species, the auditory and visual perceptual systems can identify stimuli that often are far from the body, which from an adaptive point of view might provide the advantage of recognizing potentially dangerous events when they are still at a safe distance. Hence, it would not be surprising if the attentional system had evolved to be coupled longer to auditory and visual stimuli than to tactile stimuli. In addition, once a given somatosensory stimulation (such as the touch of the finger-pads in the present study) has been identified as a non-dangerous event, it might become useless to keep attention linked to the tactile stimulation for a long time. Also note that vision and audition may hold attention longer than taction for different reasons. The visual stimulus is the slowest of the three to reach the cortex, with a latency twice as great as the fastest auditory or tactile responses (e.g., Bridgeman 1988). Thus a longer period of compulsory attentional capture would be necessary to achieve sensory integration. The auditory system, though fast, requires a longer time to define the stimulus because the stimulus can be specified only over time, as noted in the discussion of Experiment 2. Finally, it should be noted that tactile stimuli were different in two other aspects compared to visual and auditory stimuli. First, although we claimed that in our paradigm no spatial shifts of attention occurred, S_1 and S_2 in the tactile modality were delivered either in the same or in a different finger (i.e., to different, though very close, spatial locations). Second, the request of responding with the finger contralateral to the stimulated one might have made responses to all tactile S_2 easier than to auditory and visual S_2. Even if we believe that these two factors had influenced the results very little, we cannot exclude the possibility that they might have played a role at all.

To summarize, the main result of the present study is that the onset of an irrelevant stimulus in a given modality produces an automatic deployment of attention to the corresponding modality. Such modality-driven attentional capture is revealed by a short-latency facilitation in processing a following target-stimulus when the target is ipsimodal as compared to when it is crossmodal.

These findings, showing a crossmodal attentional limitation, seem to support the hypothesis of a supramodal attentional mechanism, at least when one refers to those central processes such as stimulus identification and/or response selection (Pashler 1994). Were each modality equipped with a separate and dedicated attentional system, one should not expect any difference in performance between ipsimodal trials and crossmodal trials, which is instead expected if a single supramodal attentional system had to allocate resources from one modality to another in crossmodal trials compared to ipsimodal trials. Accordingly, recent studies on the attentional blink (AB) phenomenon provided converging evidence that this form of central attentional limitation, which was originally shown within the visual modality only, can be observed even when the first- and the second-target stimuli are presented in different sensory modalities (Arnell and Jolicoeur 1999; Dell'Acqua et al. 2001).

Whether different modalities rely on separate attentional systems or on a common supramodal system is a central issue for studies investigating crossmodal links in spatial attention (Spence 2001). In principle, attention could be totally modality independent (i.e., the strong version of the supramodal system hypothesis), totally modality dependent (i.e., multiple attentional systems for each sensory modality) or, as suggested by Driver and Spence (1998), the different modalities might have different attentional systems which, however, are not fully independent and converge to some extent (i.e., the separable-but-linked modality-specific attentional systems hypothesis). Evidence is somewhat mixed in the literature, with some studies supporting the supramodal hypothesis (e.g., Eimer and Van Velzen 2002; Farah et al. 1989; McDonald and Ward 1999; Ward 1994), whereas others support the separate-but-linked hypothesis (e.g., Spence and Driver 1996; Spence et al. 2000b).

Although the present findings, as well as those from the crossmodal AB, provide evidence that seems more consistent with the supramodal hypothesis, a potentially relevant difference between these studies and the crossmodal spatial-attention studies should be noted. Whereas the present work (also see Turatto et al. 2002) and those studies that investigated crossmodal AB (Arnell and Jolicoeur 1999; Dell'Acqua et al. 2001) or AB-like effects (Jolicoeur 1999) were mainly concerned with central (i.e., post-perceptual and pre-motor) attention mechanisms, the typical crossmodal spatial attention studies were aimed at investigating how the attention moves in space (Spence 2001). In other words, there are reasons to suspect that the two different lines of research might rely on paradigms that tap distinct attentional processes. Following the distinction proposed by Johnston et al. (1995), one possibility is that crossmodal spatial attention is more closely linked to what has been termed 'input' attention, whereas crossmodal non-spatial attention (like that studied here) is more closely linked to what has been termed ‘central’ attention. Further research is needed on both ‘central’ and ‘input’ attention to provide a more direct test of the three hypotheses concerning the debate on modality-specific vs. supramodal attention systems. However, because attention seems to operate with different mechanisms at early stages compared to later stages of information processing, it might well be possible that attention selects incoming information according to either a supramodal or a modality specific system depending on the processing stage that is considered.

References

Arnell KM, Jolicoeur P (1999) The attentional blink across stimulus modalities: evidence for central processing limitation. J Exp Psychol Hum Percept Perform 25:630–648

Bridgeman B (1988) The biology of behavior and mind. Wiley, Oxford

Caclin A, Soto-Faraco S, Kingstone A, Spence C (2002) Tactile “capture” of audition. Percept Psychophys 64:616–630

Clark HH, Fox Tree J (2002) Using uh and um in spontaneous speaking. Cognition 84:73–111

Corbetta M, Miezen FM, Dobmeyer S, Shulman GL, Petersen SE (1991) Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. J Neurosci 11:2383–2402

Dell’Acqua R, Turatto M, Jolicoeur P (2001) Crossmodal attentional deficits in processing tactile stimulation. Percept Psychophys 63:777–789

Driver J, Spence C (1994) Spatial synergies between auditory and visual attention. In: Umiltà C, Moscovitch M (eds) Attention and performance XV: conscious and nonconscious information processing. MIT Press, Cambridge, MA, pp 311–331

Driver J, Spence C (1998) Attention and the crossmodal construction of space. Trends Cog Sci 7:254–262

Driver J, Spence C (2000) Multisensory perception: beyond modularity and convergence. Curr Biol 10:R731–R735

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

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 MJ, Wong AB, Monheit MA, Morrow LA (1989) Parietal lobe mechanisms of spatial attention: modality-specific or supramodal? Neuropsychologia 27:461–470

Harvey N (1980) Non-informative effects of stimuli functioning as cues. Q J Exp Psychol A 32:413–425

Johnston JC, McCann RS, Remington RW (1995) Chronometric evidence for two types of attention. Psychol Sci 6:365–369

Jolicoeur P (1999) Restricted attentional capacity between sensory modalities. Psychon Bull Rev 6:87–92

Jonides J (1981) Voluntary versus automatic control over the mind’s eye’s movement. In: Long JB, Baddeley AD (eds) Attention and performance IX. Erlbaum, Hillsdale, NJ, pp 187–203

Kawashima R, O’Sullivan BT, Roland PE (1995) Positron-emission tomography studies of cross-modality inhibition in selective attentional tasks: closing the “mind’s eye”. Proc Natl Acad Sci U S A 92:5969–5972

Kennett S, Eimer M, Spence C, Driver J (2001) Tactile-visual links in exogenous spatial attention under different postures: convergent evidence from psychophysics and ERPs. J Cog Neurosci 13:462–468

Klein RM (1977) Attention and visual dominance: a chronometric analysis. J Exp Psychol Hum Percept Perform 3:365–378

Logan GD (1980) Attention and automaticity in Stroop and priming tasks: theory and data. Cognit Psychol 12:523–553

McDonald JJ, Ward LM (2000) Involuntary listening aids seeing: evidence from human electrophysiology. Psychol Sci 11:167–171

Müller HJ, Rabbitt PMA (1989) Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. J Exp Psychol Hum Percept Perform 15:315–330

Pashler H (1994) Dual-task interference in simple tasks: data and theory. Psychol Bull 116:19–45

Posner MI (1980) Orienting of attention. Q J Exp Psychol A 32:3–25

Posner MI, Petersen SE (1990) The attention system of the human brain. Annu Rev Neurosci 13:25–42

Posner MI, Nissen MJ, Klein RM (1976) Visual dominance: an information-processing account of its origins and significance. Psychol Rev 83:157–171

Posner MI, Snyder CRR, Davidson BJ (1980) Attention and the detection of signals. J Exp Psychol Gen 109:160–174

Rogers RD, Monsell S (1995) Costs of predictable switch between simple cognitive tasks. J Exp Psychol Gen 124:207–231

Schmitt M, Postma A, de Haan E (2000) Interactions between exogenous auditory and visual spatial attention. Q J Exp Psychol A 53:105–130

Spence C (2001) Crossmodal attentional capture: a controversy resolved? In: Folk CL, Gibson BS (eds) Attraction, distraction, and action: multiple perspectives on attentional capture. North Holland, Amsterdam, pp 231–262

Spence C (2002) Multisensory attention and tactile information processing. Behav Brain Res 135:57–64

Spence C, Driver J (1996) Audiovisual links in endogenous covert spatial attention. J Exp Psychol Hum Percept Perform 22:1005–1030

Spence C, Driver J (1997a) Audiovisual links in exogenous covert spatial orienting. Percept Psychophys 59:1–22

Spence C, Driver J (1997b) On measuring selective attention to an expected sensory modality. Percept Psychophys 59:384–403

Spence C, Driver J (2000) Attracting attention to the illusory location of a sound: reflexive crossmodal orienting and ventriloquism. Neuroreport 11:2057–2061

Spence C, Lloyd D, McGlone F, Nicholls MER, Driver J (2000a) Inhibition of return is supramodal: a demonstration between all possible pairings of vision, touch, and audition. Exp Brain Res 134:42–48

Spence C, Pavani F, Driver J (2000b) Crossmodal links between vision and touch in covert endogenous spatial attention. J Exp Psychol Hum Percept Perform 26:1298–1319

Stein BE, Wallace MT, Meredith MA (1995) Neural mechanisms mediating attention and orientation to multisensory cues. In: Gazzaniga MS (ed) The cognitive neurosciences. MIT Press, Cambridge, MA, pp 683–702

Stein BE, London N, Wilkinson LK, Price DP (1996) Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. J Cog Neurosci 8:497–506

Tassinari G, Berlucchi G (1995) Covert orienting to non-informative cues: reaction-time studies. Behav Brain Res 71:101–112

Tassinari G, Campara D (1996) Consequences of covert orienting to non-informative stimuli of different modalities: a unitary mechanism? Neuropsychologia 34:235–245

Turatto M, Benso F, Galfano G, Umiltà C (2002) Nonspatial attentional shifts between audition and vision. J Exp Psychol Hum Percept Perform 28:628–639

Van Selst M, Jolicoeur P (1994) A solution to the effect of sample size on outlier elimination. Q J Exp Psychol A 47:631–650

Ward LM (1994) Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. Can J Exp Psychol 48:242–259

Ward LM, McDonald JJ, Golestani N (1998) Crossmodal control of attention shifts. In: Wright RD (ed) Visual attention. Oxford University Press, New York, NY, pp 232–268

Ward LM, McDonald JJ, Lin D (2000) On asymmetries in crossmodal spatial attention orienting. Percept Psychophys 62:1258–1264