

Exp Brain Res (2009) 195:593–602
DOI 10.1007/s00221-009-1829-y

RESEARCH ARTICLE

Competition between auditory and visual spatial cues during visual task performance

Thomas Koelewijn · Adelbert Bronkhorst ·
Jan Theeuwes

Received: 24 February 2009 / Accepted: 23 April 2009 / Published online: 13 May 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract There is debate in the crossmodal cueing literature as to whether capture of visual attention by means of sound is a fully automatic process. Recent studies show that when visual attention is endogenously focused sound still captures attention. The current study investigated whether there is interaction between exogenous auditory and visual capture. Participants performed an orthogonal cueing task, in which, the visual target was preceded by both a peripheral visual and auditory cue. When both cues were presented at chance level, visual and auditory capture was observed. However, when the validity of the visual cue was increased to 80% only visual capture and no auditory capture was observed. Furthermore, a highly predictive (80% valid) auditory cue was not able to prevent visual capture. These results demonstrate that crossmodal auditory capture does not occur when a competing predictive visual event is presented and is therefore not a fully automatic process.

Keywords Crossmodal · Attentional capture · Spatial cues · Exogenous

Introduction

It is well known that our attention may be captured by sudden visual or auditory events even when they are irrelevant for our current task. When hearing a sound or seeing a

flash, we have the tendency to direct our gaze to the location of the visual or auditory event. This enables us to respond more accurately and more quickly to events that occur at that location (Posner et al. 1980). This bottom-up or exogenous capture of attention can occur in an overt manner by making eye movements (Theeuwes et al. 1998) or in a covert manner without making eye movements (Theeuwes 1994). Within the visual domain, exogenous capture of covert attention is most often studied by means of a cueing task, in which, a localizable onset is presented at a valid or invalid target location prior to the presentation of the target. People respond faster and more accurate to validly cued targets than to invalidly cued targets. Importantly, this cueing effect occurs when the cue is valid at chance level, which indicates that it is an automatic process (e.g. Jonides 1981; Yantis and Jonides 1984). Previous research shows similar cueing effects in modalities other than vision, such as the tactile (Posner 1978; Spence and McGlone 2001) and auditory domains (Spence and Driver 1994).

Exogenous cueing effects are also known to occur across modalities. In a seminal study by Spence and Driver (1997), participants had to perform a cueing task, in which, they made an elevation judgment regarding auditory or visual targets presented to the left or right of fixation. For the elevation judgment task, the target was presented either at an “up” location above the vertical meridian or at a “down” location below the vertical meridian. This resulted in a total of four target locations with two (up and down) on each side of fixation. Cues were presented along the vertical meridian between the up and down locations on the left or right side. Therefore, only the side at which the target was presented was cued, but not its exact location. This task has become known as the orthogonal cueing task because the response dimension (up or down) is orthogonal to the cue

T. Koelewijn (✉) · A. Bronkhorst · J. Theeuwes
Cognitive Psychology, Vrije Universiteit,
van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands
e-mail: t.koelewijn@psy.vu.nl

A. Bronkhorst
TNO Human Factors, Soesterberg, The Netherlands

dimension (left or right). This has the benefit that possible response priming effects are canceled out. In Spence and Driver's (1997) study, the auditory target stimuli were generated by loudspeakers, and the visual target stimuli were generated by LED's mounted directly in front of these loudspeakers. The side at which the target could be presented was cued at chance level by either a visual or an auditory event. Results showed unimodal cueing effects in both visual and auditory domains and a crossmodal cueing effect when a visual target was preceded by an auditory cue.

An important question to address is whether attentional capture by sound as shown by crossmodal cueing is truly automatic. In order for auditory capture to be considered an automatic process, it should not be affected by top-down control (Posner 1978; Jonides 1981; Yantis and Jonides 1990). It is known that one can direct attention in a top-down (or endogenous) way to a location in space (Posner et al. 1980; Broadbent 1982). In a typical paradigm, a centrally presented arrow points to the likely target location with a high probability (e.g. 80%). Similar to exogenous cueing, endogenous cueing effects show faster responses to validly cued target locations than to invalidly cued target locations. The question whether auditory capture is sensitive to any top-down settings is addressed by several recent studies (van der Lubbe and Postma 2005; Mazza et al. 2007; Santangelo and Spence 2007; Koelewijn et al. 2009).

Van der Lubbe and Postma (2005) used a combination of endogenous and exogenous cues to investigate whether top-down control could affect attentional capture. In their study, participants performed a variation of the orthogonal cueing task, in which they had to indicate whether an arrowhead presented to the left or to the right of fixation was pointing up or down. The exogenous cue consisted of either a visual or an auditory onset and was presented 200 ms before the target (i.e. the arrowhead). Eight hundred milliseconds prior to this, cue an endogenous central cue consisting of an arrow indicated the target location with 100% validity. Note that, in this study, only elevation judgments of visual targets were made. The results showed that even though the target location was known in advance, exogenous cueing effects still occurred both in the unimodal condition and in the crossmodal condition. A recent study by Mazza et al. (2007) showed similar results. In this study, a similar design to that of Spence and Driver (1997) was used with the exception that the side at which the target would be presented was blocked. Therefore, participants knew where the target would appear and could keep their attention endogenously focused on one of the sides during an entire block. In line with the results of Van der Lubbe and Postma (2005), Mazza et al. (2007) found an auditory cueing effect on a visual target even though the target side was known to the participant. Finally, Koelewijn et al. (2009) also showed that auditory crossmodal cueing effects

occur during focused visual attention. In this study, participants performed an orthogonal cueing task in which each trial started with an endogenous cue in the form of a centrally presented arrowhead indicating the side at which the target would be presented. In Experiment 4, this arrowhead was 100% valid and placeholders indicated the possible target locations during the entire trial. In addition to the standard spatial cues, a spatially neutral auditory cue was introduced. This neutral cue had the same temporal onset as the spatial cues but was spatially diffuse. Compared to this neutral baseline condition, the results showed attentional capture by sound in the form of costs when the auditory cue was presented at the invalid location. However, no benefits were found when the target location was validly cued.

So far, several studies have demonstrated that top-down control of attention does not affect auditory capture (van der Lubbe and Postma 2005; Mazza et al. 2007; Koelewijn et al. 2009). However, no study has addressed whether auditory capture can be affected by bottom-up processes. The current study investigated auditory capture when at the same time a visual event was presented. From studies on visual search (Desimone and Duncan 1995), it is known that competing visual events share attentional resources. The more events are shown the less salient individual events become, in which, case less attention is drawn to each individual event. The present study addressed whether the bottom-up salience of a visual and auditory event is affected when these are presented at the same time.

There are previous studies that have used bimodal exogenous cues to investigate crossmodal integration (e.g. Ward 1994; Santangelo et al. 2006). Santangelo et al. 2006 investigated whether a bimodal audiovisual cue shows a super-additive effect in respect to the unimodal visual and auditory cueing effects. A super-additive effect for the bimodal cue would indicate crossmodal integration. In an orthogonal cueing task, visual, auditory, or bimodal audiovisual cues were presented prior to the presentation of a visual target. The results show similar effects sizes for the visual, auditory, and bimodal audiovisual cues, which indicates that crossmodal integration is not reflected by cueing effects. Similar results were obtained in later studies (e.g. Santangelo et al. 2008b) that show electrophysiological but no behavioral integration effects. In the bimodal condition of the Santangelo et al. 2006 study, both the visual and the auditory cue were always presented at the same location and never at opposite locations. Therefore, these results do not reveal how each individual cue contributes to the overall bimodal cueing effect. In order to show these individual contributions, the validity of each cue should be manipulated individually. Such an experiment was performed by Ward (1994) but his results are inconclusive because the paradigm used failed to show crossmodal auditory capture when unimodal auditory cues were presented. By investigating

whether the individual contributions of visual and auditory exogenous cues interact, we will learn more about the automaticity of auditory capture.

In Experiment 1, we tested how nonpredictive peripheral auditory and visual cues interact with one another and how they each influence performance in a visual spatial discrimination task. In Experiments 2 and 3, the validity of the visual or auditory cue, respectively, was raised to 80%. We wanted to determine how this top-down bias would influence the competition between the auditory and visual events over attentional resources. We hypothesized that it should be possible to prevent auditory capture of visual attention by presenting competing visual events.

Experiment 1

The paradigm used in this experiment was similar to the orthogonal cueing task used by Spence and Driver (1997). However, only elevation judgments of visual targets were made and both unimodal and bimodal cues were used. We used a within subject design consisting of experimental and control conditions. In the unimodal control conditions, both types of cues were presented separately and in the bimodal experimental conditions, auditory and visual cues were presented together. In the experimental conditions, the visual and auditory cues appeared simultaneously at the same or at opposite locations. The visual cues and targets were presented on a computer screen instead of using LED's. The loudspeakers that generated the auditory cues were located to the left and right of the computer screen.

Methods

Participants

Ten students of the Vrije Universiteit Amsterdam (3 male, mean age 21.0, age between 18 and 26) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing. Participants were informed beforehand about the experimental procedure and were naïve as to the purpose of the experiment.

Apparatus and design

Participants were seated in a dimly lit room at approximately 80 cm distance from the computer screen (17 inch, 120 Hz). The experiment was run in E-Prime 1.2 (1.2.1.847). The loudspeakers were placed at an angle of 18.3° from fixation and were aligned to the vertical middle of the screen. In the control conditions, either a visual or an auditory cue was presented that was either valid or invalid with respect to the target location. This resulted in four

conditions. The cues were presented 150 ms prior to the onset of the target. The control conditions were presented in 9 blocks containing 16 trials each. The first block was for practice purposes only, which resulted in 32 trials for each of the control conditions. In the bimodal experimental conditions, both a visual and an auditory cue were presented 150 ms prior to the onset of the target. There were four conditions, in which, the validity of the cueing side was manipulated independently for the visual and auditory cues. The experiment consisted of 9 blocks containing 16 trials each. The first block was a practice block so that 32 trials for each of the experimental conditions remained. Participants performed both the control and the experimental conditions in two separate blocks and the order of these blocks was counterbalanced over participants.

Procedure and stimuli

Figure 1 gives an example of a typical trial. At the beginning, a white fixation dot (diameter 0.2°) appeared on a black background and stayed on screen until a response was made. Participants were instructed to fixate this dot during the entire trial and to refrain from making eye movements. After a random delay time of 900–1,150 ms, the auditory and visual cue were presented. In the control conditions, this was either an auditory or a visual cue. In the experimental conditions, both an auditory and a visual cue were presented. After an SOA of 150 ms, the target appeared. The auditory cue consisting of a white noise burst that was presented for 100 ms. This cue was equiprobably presented from the left or right loudspeaker and was valid or invalid with respect to the target location. The visual cue consisted of a dark gray horizontal bar (width 0.6° and height 0.2°) presented at 11.7° left or right of fixation and was valid or invalid with respect to the target location. To prevent attention from being drawn to its offset (Theeuwes 1991), the visual cue remained on screen until a response was made. The auditory cue was switched off because there is no evidence that auditory offsets are able to draw attention.

The visual target consisted of a white dot (diameter 0.2°) that was presented for 140 ms. During the entire trial, placeholders were displayed on both sides of the screen indicating the possible target locations. This allowed participants to focus their attention to a predefined location. The placeholders were thin light gray lined squares with a width and height of 1.3° that indicated the area in which targets could appear. On each side of the screen, two connecting placeholders were shown—one for targets displayed above the vertical meridian of the screen, and the other for targets below the vertical meridian. The horizontal centre of the squares was separated by an angle of 10.5° from the centre of the screen. The two target locations at each side were positioned above each other on average 0.6° above and

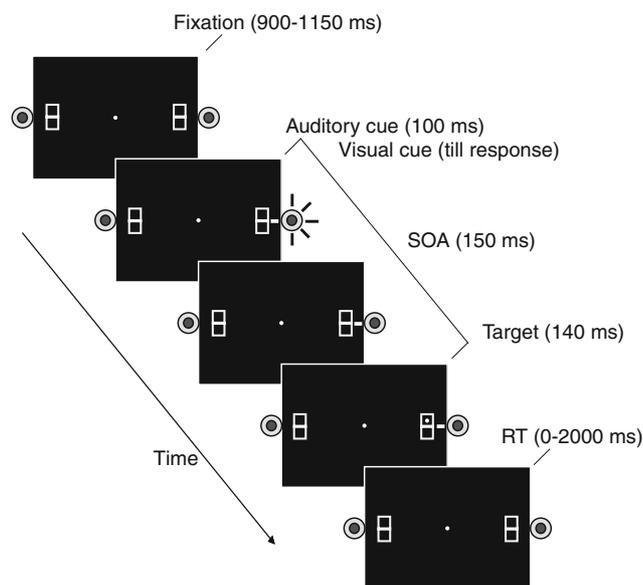


Fig. 1 Schematic representation of the paradigm used. Participants performed an orthogonal cueing task where they had to discriminate between targets presented in the upper or lower square or the placeholder. Targets were presented on the left or right side of the screen and were preceded independently with an SOA of 150 by a nonpredictive visual or auditory cue. In the example given, both the auditory and the visual cue are valid

below the vertical meridian of the screen. The exact location was varied by placing the targets randomly within a range of 0.3° from the centre of the placeholder. The target appeared at one of the four locations at chance level.

The participants' task was to report in a speeded but accurate fashion whether the target appeared above or below the vertical meridian of the screen. They did this using the number pad of a QWERTY keyboard, by pressing the numbers 8 or 2 with their right and left index finger, respectively. Because of the task's orthogonal design, there was no need to balance out for possible motor response effects. Responses had to be made within a time window of 2,000 ms after target onset. After the response, participants received feedback when they had made an error—the fixation dot then turned red for 150 ms. After each trial, an interval of 850 ms followed before starting the next trial. Following each block, participants received further feedback in the form of a percentage correct score and a mean reaction time. At the beginning of the experiment, participants were told that both auditory and visual cues would provide no information about the location of the targets and therefore could be ignored.

Results

Control conditions

For each subject, the average score and its standard deviation were calculated for each condition over all data.

Table 1 Results of the control condition of Experiment 1, 2, and 3

Exogenous cue modality	Exogenous cue validity		Cueing effect	
	Valid	Invalid		
Experiment 1				
Auditory	RT	395 (41)	409 (49)	14
	%	1.59	2.57	
Visual	RT	420 (51)	464 (66)	44
	%	5.35	7.36	
Experiment 2				
Auditory	RT	361 (56)	378 (60)	17
	%	8.47	9.23	
Visual	RT	362 (47)	425 (63)	63
	%	9.13	14.18	
Experiment 3				
Auditory	RT	379 (20)	407 (29)	28
	%	2.26	4.20	
Visual	RT	412 (27)	463 (24)	51
	%	3.78	4.35	

Shown are mean reaction times (RTs; in milliseconds), their standard deviations (in parentheses), and percentages of errors, for visual targets as a function of exogenous cue modality (auditory and visual), exogenous cue validity (valid and invalid), together with the cueing effects for each modality

Reaction times above or below average by 2.5 times their standard deviation were considered to be outliers (in total 2.1%) and were removed from further analysis. This method of determining outliers was used for all data analyses in this study. For the remaining trials, mean reaction times for the correct response trials (95.8%) were calculated for each condition. An overview of the mean reaction times, their standard deviations, and the mean error scores, for each condition and for all experiments, is shown in Table 1. An ANOVA was performed on RT with cue validity (valid and invalid) and cue modality (auditory and visual) as factors. The outcome revealed a significant effect for cue validity [$F(1,9) = 15.077$, $MSE = 552.058$, $p < 0.005$] and for cue modality [$F(1,9) = 14.152$, $MSE = 1161.858$, $p < 0.005$]. In addition, the results show an interaction between cue validity and cue modality [$F(1,9) = 10.560$, $MSE = 228.969$, $p < 0.05$] indicating a stronger cueing effect for visual cues than for auditory cues. Two pairwise two-tailed t tests for the separate cue modalities were conducted. These results show a cueing effect for both the visual ($p = 0.004$) and auditory ($p = 0.018$) cues.

A similar ANOVA conducted on error data revealed no effect, which indicates that there was no speed accuracy tradeoff.

Experimental conditions

Outliers (in total 2.0%) were removed from further analysis. For the remaining trials, mean reaction times for the correct response trials (94.0%) were calculated for each subject for each condition. The mean reaction times for each condition averaged over subjects are plotted in Fig. 2. The error bars in all figures represent the 0.95 confidence interval for the auditory cueing main effect, following Loftus and Masson (1994). An ANOVA was conducted on RT with as factors auditory cueing and visual cueing. The results show an effect for auditory cueing [$F(1,9) = 6.424$, $MSE = 343.281$, $p < 0.05$] and for visual cueing [$F(1,9) = 7.377$, $MSE = 4396.303$, $p < 0.05$]. No interaction between auditory and visual cueing ($F < 1$) was observed.

A similar ANOVA conducted on error data revealed a trend for visual cueing [$F(1,9) = 4.506$, $MSE = 0.007$, $p = 0.063$] indicating that participants made more errors on invalid visual cueing trials (14.2%) than on valid trials (9.1%). The error data indicates that there was no speed accuracy tradeoff.

Cueing effects

Two paired sample *t* test were conducted to test for possible size differences of auditory and visual cueing effects between the control and experimental conditions. No difference between the unimodal auditory cueing effect and the experimental auditory main cueing effect ($t < 1$), and between the unimodal visual cueing effect and the experimental visual main cueing effect ($t < 1$) were observed.

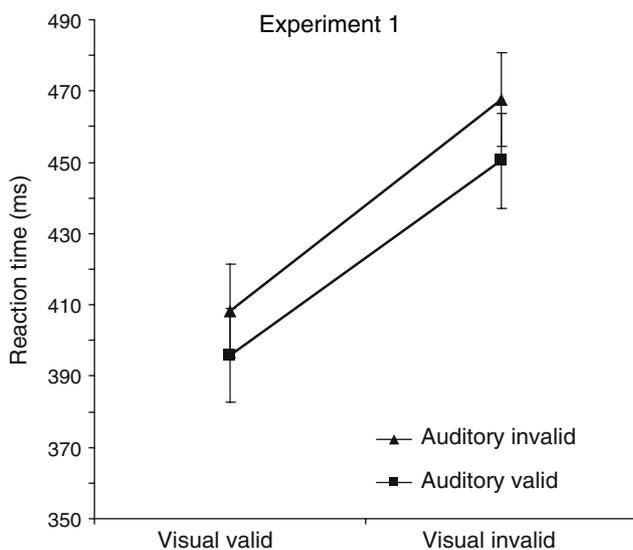


Fig. 2 Graph with the results of the experimental part of Experiment 1. The graph shows the average reaction time (ms) for all auditory cue (valid and invalid) and visual cue (valid and invalid) validity combinations. The error bars show the 0.95 confidence intervals for the auditory cueing main effect (Loftus and Masson 1994)

Discussion

The results show auditory as well as visual exogenous cueing effects in both the unimodal control conditions and the bimodal experimental conditions. In addition, no differences in cueing effect size between the control and experimental conditions were observed. This indicated that the overall bimodal cueing effect is additive and made up of the summation of visual and auditory cueing effects. The current results indicate that auditory capture is not affected by a competing exogenous visual cue. Earlier studies (van der Lubbe and Postma 2005; Mazza et al. 2007; Koelewijn et al. 2009) show that auditory capture is not affected by top-down control. The current results suggest that the same holds for bottom-up competition. Overall, these outcomes seem to indicate that auditory capture is an automatic process not affected by other competing processes.

Note, however, that this conclusion is inconsistent with that of Santangelo and Spence (2008) who showed no auditory capture when attention was endogenously focused by an RSVP stream. In their study, participants performed an RSVP task at the center of the display while exogenous cues were presented in the periphery. In other studies (van der Lubbe and Postma 2005; Mazza et al. 2007; Koelewijn et al. 2009), information about the upcoming target location was presented prior to the presentation of both cue and target. For example, the central arrowheads in (van der Lubbe and Postma 2005; Koelewijn et al. 2009) pointed towards the target location before the target was presented. Therefore, from trial to trial, participants had to endogenously refocus their attention at the start of each trial, and were required to maintain their attention focused on the target location during a trial. The RSVP stream used by Santangelo and Spence (2008) contained most targets and was therefore inevitably presented at the most valid target location so that participants could keep their attention focused at the centre during the entire block. It could be that it is harder to disengage attention from stimuli presented at the same location as the target than from a target location indicated symbolically by means of an arrowhead. To test this assumption, in Experiment 2, predictive peripheral visual cues were presented at the same time as the exogenous auditory cues.

Experiment 2

Experiment 1 demonstrated that auditory capture still occurs when a nonpredictive visual cue is presented at the same time as the auditory cue. In Experiment 2, we investigated whether making the visual cue predictive would affect capture by the auditory cue. The validity of the visual cue was set at 80% while the validity of the auditory cue

remained at chance level. Again, we used a within-subject design with bimodal experimental and unimodal control conditions.

Methods

Participants

Ten new students of the Vrije Universiteit Amsterdam (4 male, mean age 21.4, age between 16 and 30) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing. Participants were informed beforehand about the experimental procedure and were naïve as to the purpose of the experiment.

Apparatus and design

The setup and design were basically identical to the one used in Experiment 1. However, this time, the visual cue was valid in 80% of the trials. The control conditions were presented in 9 blocks containing 28 trials each and the first block was for practice purposes. The visual cue was presented in 160 trials; in 128 trials it was valid and in the remaining 32 trials invalid. The auditory cue was presented in 64 trials and was valid in 50% of the trials. The experimental conditions were presented in 9 blocks containing 40 trials each and also here the first block was for practice purposes. The 2 visual valid conditions (visual valid–auditory valid, visual valid–auditory invalid) consisted of 128 trials each, and the 2 visual invalid conditions (visual invalid–auditory valid, visual invalid–auditory invalid) consisted 32 trials each. Again, participants performed both the control and the experimental conditions in separate blocks and the order of these blocks was counterbalanced over participants.

Procedure and stimuli

Procedure and stimuli were mostly identical to the one used in the previous experiment. However, at the beginning of the experiment, participants were now told that the visual cues would provide information about the location of the targets and that they should make use of this information. The auditory cue would provide no information about the location of the targets and therefore could be ignored.

Results

Control conditions

Outliers (in total 2.7%) were removed from further analysis. For the remaining trials, mean reaction times for the

correct response trials (90.5%) were calculated for each condition. An ANOVA was performed on RT with cue validity (valid and invalid) and cue modality (auditory and visual) as factors. The outcome revealed a significant effect for cue validity [$F(1,9) = 58.023$, $MSE = 270.267$, $p < 0.001$] and for cue modality [$F(1,9) = 15.432$, $MSE = 367.067$, $p < 0.005$]. In addition, the results show an interaction between cue validity and cue modality [$F(1,9) = 12.571$, $MSE = 413.511$, $p < 0.01$] indicating a stronger cueing effect for visual cues than for auditory cues. Two pairwise two-tailed t tests for the separate cue modalities were conducted. These results show a cueing effect for both the visual ($p < 0.001$) and auditory ($p = 0.001$) cues.

A similar ANOVA conducted on error data revealed no effect, which indicates that there was no speed accuracy tradeoff.

Experimental conditions

Outliers (in total 2.3%) were removed from further analysis. For the remaining trials, mean reaction times for the correct response trials (93.0%) were calculated for each subject for each condition. The mean reaction times for each condition averaged over subjects are plotted Fig. 3. An ANOVA was conducted on RT with as factors auditory cueing and visual cueing. The results show an effect for visual cueing [$F(1,9) = 352.334$, $MSE = 129.892$, $p < 0.001$]. Neither auditory cueing effect ($F < 1$) nor an interaction between auditory and visual cueing ($F < 1$) was observed.

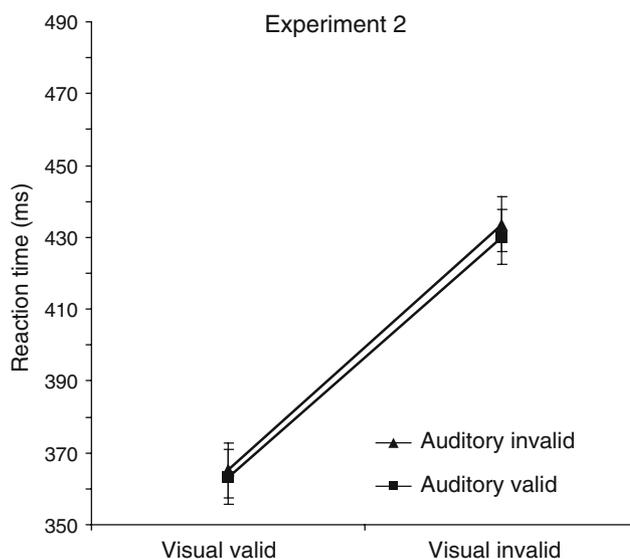


Fig. 3 Graph with the results of the experimental part of Experiment 2 showing average reaction time (ms) for all auditory cue (valid and invalid) and visual cue (valid and invalid) validity combinations. The error bars show the 0.95 confidence intervals for the auditory cueing main effect

A similar ANOVA conducted on error data revealed no effects, which indicate that there was no speed accuracy tradeoff.

Discussion

The results of this experiment show no exogenous auditory cueing effect when the auditory cue was presented together with an 80% valid visual cue. When comparing the current results to those of Experiment 1, it appears that the validity of the visual cue is an important factor determining the occurrence auditory capture. Interestingly, we do not know whether the same will hold for visual cueing. In other words, are we able to prevent visual capture when we are presenting a nonpredictive visual cue at the same time as highly predictive auditory cue. To test whether such symmetry in competition between modalities exist, we conducted a third experiment.

Experiment 3

Experiment 2 demonstrated that auditory capture does not occur when a predictive visual cue is presented at the same time as the auditory cue. In Experiment 3, we tested whether a predictive auditory cue could affect visual capture. Therefore, the validity of the auditory cue was set at 80% while the validity of the visual cue now remained at chance level. Again, we used a within-subject design with a bimodal experimental and unimodal control condition.

Methods

Participants

Ten new students of the Vrije Universiteit Amsterdam (5 male, mean age 23.8, age between 18 and 44) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing. Participants were informed beforehand about the experimental procedure and were naïve as to the purpose of the experiment.

Apparatus and design

The setup and design were basically the same to the one used in Experiment 2. However, this time, the auditory cue was valid in 80% of the trials while the visual cue was valid in 50% of the trials. Again, participants performed both the control and the experimental conditions in separate blocks and the order of these blocks was counterbalanced over participants.

Procedure and stimuli

Procedure and stimuli were basically the same as used in the previous experiments. However, at the beginning of the experiment, participants were now told that the auditory cues would provide information about the location of the targets and that the visual cue could be ignored.

Results

Control conditions

Outliers (in total 1.6%) were removed from further analysis. For the remaining trials, mean reaction times for the correct response trials (97.0%) were calculated for each condition. An ANOVA was performed on RT with cue validity (valid and invalid) and cue modality (auditory and visual) as factors. The outcome revealed a significant effect for cue validity [$F(1,9) = 81.925$, $MSE = 187.567$, $p < 0.001$] and for cue modality [$F(1,9) = 120.885$, $MSE = 165.289$, $p < 0.001$]. In addition, the results show an interaction between cue validity and cue modality [$F(1,9) = 13.203$, $MSE = 98.433$, $p < 0.01$] indicating a stronger cueing effect for visual cues than for auditory cues. Two pairwise two-tailed t tests for the separate cue modalities were conducted. These results show a cueing effect for both the visual ($p < 0.001$) and auditory ($p < 0.001$) cues.

A similar ANOVA conducted on error data revealed no effects, which indicate that there was no speed accuracy tradeoff.

Experimental conditions

Outliers (in total 2.3%) were removed from further analysis. For the remaining trials, mean reaction times for the correct response trials (97.5%) were calculated for each subject for each condition. The mean reaction times for each condition averaged over subjects are plotted Fig. 4. An ANOVA was conducted on RT with as factors auditory cueing and visual cueing. The results show an effect for auditory cueing [$F(1,9) = 8.153$, $MSE = 158.003$, $p < 0.05$] and for visual cueing [$F(1,9) = 25.481$, $MSE = 684.069$, $p < 0.005$]. A trend in the interaction between auditory and visual cueing [$F(1,9) = 5.004$, $MSE = 237.692$, $p = 0.051$] was observed.

A similar ANOVA conducted on error data revealed an effect for visual cueing [$F(1,9) = 18.116$, $MSE = 0.000$, $p < 0.001$] indicating that participants made more errors on invalid visual cueing trials (1.7%) than on valid trials (3.3%). The error data indicates that there was no speed accuracy tradeoff.

Discussion

The results show auditory as well as visual exogenous cueing effects in both the unimodal control conditions and the bimodal experimental conditions. Additionally, the results show a trend for the interaction between visual and auditory cueing, which is illustrated in Fig. 4. The current results suggest that a highly predictive auditory event cannot prevent visual capture. Because it was found in Experiment 2 that a predictive visual event is able to prevent auditory capture, we find an asymmetry in how the visual and auditory modalities affect one another. This asymmetry can most likely be attributed to the modality in which the task was performed. In all three experiments, participants performed a visual spatial discrimination task. They had to discriminate between a visual target event that was presented either above or below the vertical meridian. The fact that participants had to respond to visual target events made them possibly more sensitive towards the visual cues as well. This bias towards the detection of visual events might explain why visual capture seems to be less sensitive for crossmodal competition than auditory capture.

General discussion

There is debate in the cueing literature as to whether crossmodal auditory attentional capture is fully automatic. On the one hand, there are studies that have shown that endogenous focusing of visual attention does not affect auditory

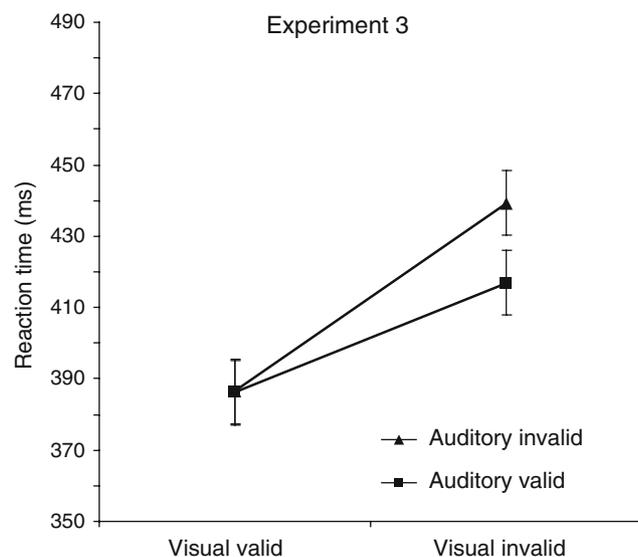


Fig. 4 Graph with the results of the experimental part of Experiment 3 showing average reaction time (ms) for all auditory cue (valid and invalid) and visual cue (valid and invalid) validity combinations. The error bars show the 0.95 confidence intervals for the auditory cueing main effect

capture (van der Lubbe and Postma 2005; Mazza et al. 2007; Koelewijn et al. 2009). On the other hand, Santangelo and Spence (2007) have recently shown that when visual attention is focused by means of an additional task auditory attentional capture ceases to exist. The current study investigated what happens to auditory capture when at the same time a peripheral visual event is presented. In Experiment 1, nonpredictive peripheral visual cue was presented to test whether bottom-up competition would affect auditory capture. In Experiment 2, a predictive peripheral visual cue was presented to test whether making a cue more endogenous could explain the discrepancy between previous studies. Additionally, in Experiment 3, we tested whether a predictive auditory cue could prevent capture by a nonpredictive visual cue.

The results of the control condition of Experiment 1 replicated the findings of Spence and Driver (1997) by showing cueing effects for both visual and auditory unimodal cues on a visual elevation judgement task. In addition, Experiment 1 showed visual and auditory cueing effects when both auditory and visual cues were present in one trial. This means that auditory capture was not affected by the nonpredictive visual event. Whereas in Experiment 1, all cues were nonpredictive, in Experiment 2, the visual cue was 80% valid but the auditory cue remained nonpredictive. Here, the results showed no auditory cueing effect when the auditory and visual events co-occurred, while a clear unimodal auditory cueing effect was still observed. These results indicate that capture of visual attention by means of an auditory event does not occur when a predictive peripheral visual event is presented at the same time. In Experiment 3, the auditory cue was 80% valid and the visual cue was nonpredictive. Here, the results show both auditory and visual cueing effect. In other words, there seems to be an asymmetry in how vision and sound affect one another.

The pattern of results observed in Experiment 2 is similar to those reported by Santangelo and Spence (2007), in that both studies show that auditory capture of visual attention can be prevented. However, in the study by Santangelo and Spence (2007), visual attention was focused by means of an RSVP stream presented at visual fixation. The authors suggest that this RSVP stream induced high-perceptual load and that this load prevented auditory capture. In contrast, in the current study, visual events were simple onsets presented in the periphery and therefore unlikely to create a high perceptual load. Therefore, it seems that a high-perceptual load is not a prerequisite for suppression of auditory capture. Both the current results and the results by Santangelo and Spence (2007) suggest that auditory capture is not a fully automatic process.

It is interesting to compare the current results with those of Koelewijn et al. (2009). In both studies, information

about the probable location of the target was provided using top–down information. In addition, the auditory cues as well as the placeholders that were used to demarcate where target stimuli could occur were identical. Nevertheless, the outcome of both studies is fundamentally different: whereas auditory capture was still observed when an arrowhead indicated the side where the target would occur with 100% accuracy, no capture occurred when the side was indicated by an 80% predictive peripheral visual cue. One explanation of the difference could be that a cue presented close to the target location results in a smaller attentional focus than a central arrowhead pointing towards a target location. However, the comparison of the results for valid, neutral and invalid cueing in Koelewijn et al. (2009) demonstrate that attentional benefits disappear when a 100% valid arrowhead is used in combination with placeholders. Apparently, in that condition, there is already such a narrow focus that the additional (valid) auditory cue does not affect performance anymore. Possibly the use of placeholders of similar size surrounding the target locations may also have contributed to this effect. The fact remains that the endogenous system is not able to prevent for auditory capture towards an invalid target location. This suggests that it is not the size of the attentional focus that matters but whether the attentional system allows for an easy disengagement of attention.

We suggest that it may be harder to disengage attention when the target location is cued by a predictive onset at target location then by a predictive arrowhead pointing towards target location. This may explain the discrepancy between the current results and those of (Koelewijn et al. 2009). The current results show that the presence of an 80% valid, localizable visual event prevents auditory capture. This type of events recruits both bottom–up and top–down resources. Top–down, because these cues were predictive and bottom–up, because there is a peripheral onset which is similar to the pure exogenous cues used in Experiment 1. The importance of this top–down or endogenous factor in suppressing auditory capture is evident when we compare the result of Experiments 1 and 2 where auditory capture is prevented only when the visual cue is predictive. This is so even though endogenous attention alone is not sufficient to suppress auditory capture (van der Lubbe and Postma 2005; Mazza et al. 2007; Koelewijn et al. 2009). Bottom–up or exogenous resources are drawn to the onset of the visual cue which is presented at target location. As shown by the results of Experiment 2, location of this cue is of high importance. As such, it may be that the concurrent activation of both the endogenous and exogenous attentional system results in the ability to prevent auditory capture.

The results of the current study are consistent with those of Santangelo and Spence (2007) that showed no capture when attention was focused by means of an RSVP stream. Santangelo and Spence (2008) argue that attentional

capture by a peripheral onset does not occur under circumstances of high perceptual load. However, the current results show that high perceptual load is not a necessity in preventing attentional capture. It should be noted that in addition to generating high perceptual load the onset of each individual item in the RSVP stream could also draw on bottom-up attentional resources. In a recent study, Santangelo et al. (2008a) tested whether these onsets could explain their results, by replacing the RSVP stream by a central morphing shape. Because of the morphing, no abrupt visual onsets occur and therefore the shape should not draw exogenous attention. In addition, under these circumstances, no cueing effect for visual peripheral events was observed. However, just as in similar to the RSVP stream most targets were presented within this morphing shape. Therefore, the morphing shape was inevitably presented at the most valid target location. The importance of this correspondence between cue and target location is something we also find in the current study.

To conclude, the current results show that auditory capture of visual attention will not occur when a predictive peripheral visual event is presented at the same time. Therefore, it seems that both visual and auditory events compete for the same attentional resources. In order for attention to remain focused on the correct target location, a correspondence between cue and target location is of importance. Based on these outcomes, we conclude that attentional capture by sound is not a fully automatic process and can be prevented at an attentional level.

Acknowledgments This research was supported by a Dutch Technology Foundation STW grant (VET.7079), from the applied science division of NWO and the Technology Program of the Ministry of Economic Affairs to Jan Theeuwes and Adelbert Bronkhorst.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Broadbent DE (1982) Task combination and selective intake of information. *Acta Psychol* 50:253–290
- Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193–222
- Jonides J (1981) Voluntary vs. automatic control over the mind's eye's movements. In: Long JB, Baddeley AD (eds) *Attention and performance IX*. Erlbaum, Hillsdale, pp 187–203
- Koelewijn T, Bronkhorst A, Theeuwes J (2009) Auditory and visual capture during focused visual attention. *J Exp Psychol Hum Percept Perform* (in press)
- Loftus GR, Masson MEJ (1994) Using confidence-intervals in within-subject designs. *Psychon Bull Rev* 1:476–490
- Mazza V, Turatto M, Rossi M, Umiltà C (2007) How automatic are audiovisual links in exogenous spatial attention? *Neuropsychologia* 45:514–522

- Posner MI (1978) Chronometric explorations of mind. Erlbaum, Hillsdale
- Posner MI, Snyder CRR, Davidson BJ (1980) Attention and the detection of signals. *J Exp Psychol Gen* 109:160–174
- Santangelo V, Spence C (2007) Multisensory cues capture spatial attention regardless of perceptual load. *J Exp Psychol Hum Percept Perform* 33:1311–1321
- Santangelo V, Spence C (2008) Is the exogenous orienting of spatial attention truly automatic? Evidence from unimodal and multisensory studies. *Conscious Cogn* 17:989–1015
- Santangelo V, Van der Lubbe RHJ, Belardinelli MO, Postma A (2006) Spatial attention triggered by unimodal, crossmodal, and bimodal exogenous cues: a comparison of reflexive orienting mechanisms. *Exp Brain Res* 173:40–48
- Santangelo V, Finoia P, Raffone A, Belardinelli MO, Spence C (2008a) Perceptual load affects exogenous spatial orienting while working memory load does not. *Exp Brain Res* 184:371–382
- Santangelo V, Van der Lubbe RHJ, Belardinelli MO, Postma A (2008b) Multisensory integration affects ERP components elicited by exogenous cues. *Exp Brain Res* 185:269–277
- Spence C, Driver J (1994) Covert spatial orienting in audition—exogenous and endogenous mechanisms. *J Exp Psychol Hum Percept Perform* 20:555–574
- Spence C, Driver J (1997) Audiovisual links in exogenous covert spatial orienting. *Percept Psychophys* 59:1–22
- Spence C, McGlone FP (2001) Reflexive spatial orienting of tactile attention. *Exp Brain Res* 141:324–330
- Theeuwes J (1991) Exogenous and endogenous control of attention—the effect of visual onsets and offsets. *Percept Psychophys* 49:83–90
- Theeuwes J (1994) Stimulus-driven capture and attentional set—selective search for color and visual abrupt onsets. *J Exp Psychol Hum Percept Perform* 20:799–806
- Theeuwes J, Kramer AF, Hahn S, Irwin DE (1998) Our eyes do not always go where we want them to go: capture of the eyes by new objects. *Psycholog Sci* 9:379–385
- van der Lubbe RHJ, Postma A (2005) Interruption from irrelevant auditory and visual onsets even when attention is in a focused state. *Exp Brain Res* 164:464–471
- Ward LM (1994) Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual-attention. *Can J Exp Psychol* 48:242–259
- Yantis S, Jonides J (1984) Abrupt visual onsets and selective attention - evidence from visual-search. *J Exp Psychol Hum Percept Perform* 10:601–621
- Yantis S, Jonides J (1990) Abrupt visual onsets and selective attention—voluntary versus automatic allocation. *J Exp Psychol Hum Percept Perform* 16:121–134