Exp Brain Res (2013) 231:13–18
DOI 10.1007/s00221-013-3663-5

RESEARCH ARTICLE

Attention to endogenous and exogenous cues affects auditory localization

Lindsay Collins · James Schirillo

Received: 21 December 2012 / Accepted: 28 July 2013 / Published online: 14 August 2013

Abstract Three experiments examine attentional differences in auditory localization using either endogenous or exogenous visual cues. Participants were presented with visual cues on a computer screen and asked to localize auditory targets presented through headphones. In conditions in which the auditory and visual stimuli traveled in the same direction, participants showed illusory directional hearing (Hari in Neurosci Lett 189:29–30, 1995) in that the targets were perceived to travel through the head. In conditions in which the directions of the auditory and visual stimuli were conflicting, participants localized the auditory targets as traveling in the direction of the visual cues. These data suggest that visual capture plays a predominate role in the processes of auditory localization that occurs within the head. Additionally, endogenous response times were significantly greater than exogenous response times. We propose this is the result of additional time required to shift one's attentional window in the endogenous condition.

Keywords Cross-modal localization · Endogenous attention · Exogenous attention

Introduction

The multisensory literature presents conflicting ideas of how one attends to visual–auditory signals. Two prominent theories are visual capture (Hay et al. 1965) and the supramodal theory (Farah et al. 1989). Visual capture refers to a phenomenon in which an auditory target is perceived to originate from the location of a visual cue when the two signals are actually spatially disparate. Visual capture is supported in research by Spence and Driver (1996) who suggest that one can selectively attend to the visual modality without interference from the auditory modality but one cannot selectively attend to the auditory modality without interference from the visual modality. In contrast, the supramodal theory (Ward 1994) suggests that the visual and auditory modalities are equally influential in making auditory localizations.

The present study seeks to shed light on this debate by using both endogenous and exogenous visual cues while also having the auditory localizations occur within the head. Endogenous cues (used in Experiment 1) refer to visual stimuli which are only symbolic of a target location (i.e., a rotating arrow at the center of the screen pointing to a specific location), or (Experiment 3) numbers flashed in locations of placeholders, and exogenous cues (used in Experiment 2) refer to visual stimuli in which a cue is directly indicative of the target location (i.e., circles located at the tips of where the arrows would have pointed in the endogenous condition). The localizations made were of auditory clicks presented over headphones. While our general research question addresses visual capture and the supramodal theory of perception, it specifically determines any differential effects of an endogenous or exogenous visual cue on the localization of an auditory target. This study differs from the research done by Soto-Faraco et al. (2004) who found that vision could modulate auditory motion, in that the current study presents sounds that are internal to the head.

Based on previous findings (Mays and Schirillo 2005), we hypothesized that auditory localizations of sounds that were perceived within the head can be influenced by visual cues that obviously occur outside the head (the visual

L. Collins · J. Schirillo (✉)
Department of Psychology, Wake Forest University, Winston-Salem, NC 27109, USA
e-mail: schirija@wfu.edu

cues appear on a computer monitor). We predicted that this hypothesis would be revealed in two ways. First, the auditory signals would be perceived to step through the participant's head, from one side of the head to the other. This contradicts the actual locations of the auditory stimuli, which were binaurally presented clicks with inter-aural time differences so that they should have been heard in either just the far left or just the far right side of the head. Hari (1995) demonstrated this purely auditory phenomenon that she titled illusory directional hearing. Second, we expected that trials in which visual cues do not accurately coincide with the direction of auditory target change should show an effect of the location of the visual cue such that the auditory localization is biased toward the visual cue. This should occur even though the sounds are not external, like the visual cues, but within the head. We hypothesize that the differences between auditory target localizations will be less pronounced when there is conflicting auditory–visual input, but that participants will always be more likely to localize the sound in the direction of the visual cue, thus demonstrating visual capture.

Finally, we hypothesize that endogenous visual cues (i.e., arrows) will elicit longer response times than those associated with exogenous visual cues (i.e., circles) due to the time required to transfer a participant's attentional window from the center of the arrow to the location of which the cue was indicative (Chakravarthi and VanRullen 2011). These extended times were replicated in Experiment 3, which used numbers instead of arrows.

Method

Participants

All procedures were approved by the Institutional Review Board of Wake Forest University and were performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. There were 120 participants (40 in each of the three experiments; ages 18–24; 56 males and 64 females). Neither age nor gender served as a basis for selection. All participants were undergraduate students at Wake Forest University enrolled in an introductory psychology course. Participants received course credit for their participation.

Apparatus

Participants sat in a quiet, dimly lit room. Their chins were placed on a chin rest located 24 inches directly in front of a CRT screen (19″ CTX, model EX95950F, with a 75 Hz refresh rate). They wore circumaural headphones (Koss; Pro/4AA; ~50 dB A at peak measurement).

Procedure

For all experiments, each trial consisted of eight click-pairs (40 dB A weighted) heard over headphones, that is, binaural clicks with small inter-aural time differences. Each one ms click had a 0.8 ms delay between ears to create the inter-aural time differences. The side in which the first click was produced is the leading side resulting in two conditions (right-ear leading and left-ear leading). One half of the trials began with four left-ear leading clicks followed by four right-ear leading clicks, while the other half was reversed (starting with four right-ear leading clicks and then four left-ear leading clicks).

In addition to the clicks, participants were presented with visual cues located at the center of the CRT monitor (Fig. 1). The onset of the visual directional cue occurred before the onset of the auditory directional cue. This was because of the relatively slow refresh rate of the screen. This may influence why the visual directional cue dominated directional judgments. Experiment 2 used a rotating arrow at the center of the screen that flashed eight times, each time in a different orientation (endogenous attention; Fig. 1a–c) while Experiment 3 used circles at the locations of the tips of where the arrows would have been in the endogenous condition to indicate cued locations (exogenous attention; Fig. 1b–d). The orientations proceeded in one of the four directions: right-to-left above the horizontal meridian (Fig. 1a, b), right-to-left below the horizontal meridian (Fig. 1c, d), left-to-right above the horizontal meridian (Fig. 1a, b), and left-to-right below the horizontal meridian (Fig. 1c, d). The first arrow cue was presented horizontally to either the far right or left of the screen with each subsequent cue changing in orientation by 25.71° until the arrow cue was horizontally oriented at the opposite side. There were also two control conditions in which the cue flashed eight times in the vertical meridian with the arrow pointing either up or down in the endogenous condition, and in the exogenous condition, the circles were at the top or bottom where the vertical meridian arrows tips would have been in the endogenous condition.

Experiment 3 was an endogenous experiment that used numbers only. The numbers one–eight or eight–one appeared in the sequence of the arrow heads.

Each visual cue lasted 100 ms; the corresponding auditory target was presented after 50 ms (i.e., in the center of the visual cue duration; Fig. 2). Participants were instructed to fixate their attention on the visual cue. Following the presentation of the eight visual and auditory stimuli, participants were instructed to record the location from which they perceived only the eight sounds to have originated. Responses were recorded using eight computerized sliders (one corresponding to each click-pair) presented simultaneously on the CRT monitor after a given

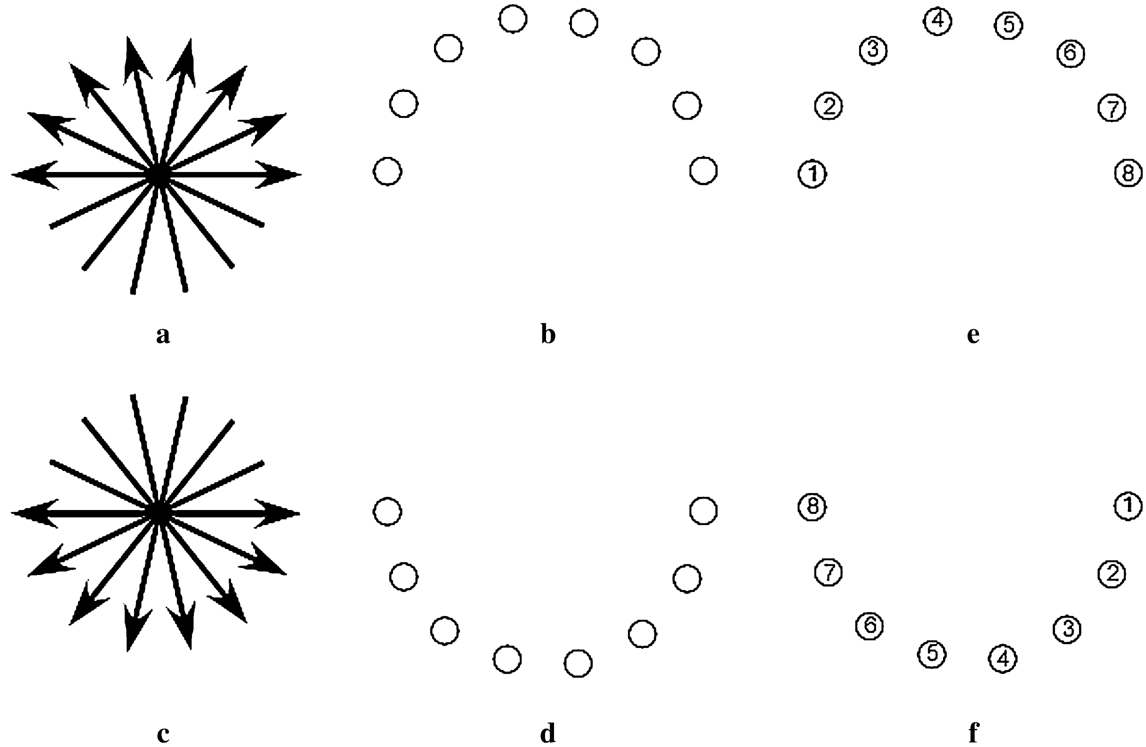

Fig. 1 Sequence of eight visual stimuli; direction stimuli flashed are not indicated. **a**, **c**, **e** and **f** Endogenous conditions, **b**–**d** exogenous conditions (not to scale)

series of eight click-pairs. The sliders were labeled −10 to +10, with −10 representing the far left and +10 representing the far right (thus 0 was meant to indicate the center of the head).

In total, for each experiment, there were twelve conditions (six visual cue orientations and two auditory conditions) per block. Each block was presented five times for a total of sixty trials per participant. Each entire experiment lasted approximately one half hour. Data obtained were graphed to determine any effects of the direction of the visual cue on the localization of the auditory target. The significance of these effects was determined by taking the difference between the fourth and fifth clicks (where there is a change from right-ear leading to left-ear leading or vice versa). A MANOVA was conducted to test this significance between conditions.

Results

Localization: endogenous (arrows)

There were six conditions in total. These included two non-conflicting conditions: left-ear leading with the visual cue beginning at the left (left–left) and right-ear leading with the visual cue beginning at the right (right–right), two conflicting conditions.: left-ear leading with the visual cue beginning at the right (left–right) and right-ear leading with the visual cue beginning at the left (right–left), and two control conditions with either right-ear leading or left-ear leading and an uninformative visual cue on the vertical meridian, which did not move. Perceived locations in the non-conflicting conditions followed the direction of the auditory target. However, instead of being perceived as four clicks from one side followed by four clicks from the other side, participants' localizations showed a relatively equal change perceived between each click. The average range for left and right-ear-leading click-pairs was 10.9 units. When compared to the control conditions with uninformative arrows, the non-conflicting signals were significantly different in the right-leading and left-leading conditions ($F = 33.761$, $df = 1, 78$, $p = .001$, $\eta^2 = .302$; see Fig. 3). The average range for left and right-ear-leading click-pairs in the control condition was 9.8 units.

In conditions with conflicting signals, this pattern was reversed. The left-ear leading condition was perceived to travel from right-to-left, and the right-ear leading condition was perceived to travel from left-to-right. As can be seen in Fig. 3, this effect was not as strong as 9.8 units ($t = 2.76$, $df = 78$, $p = .007$, $\eta^2 = .643$). The average range for left and right-ear-leading click-pairs in the conflicting condition was 6.0 units.

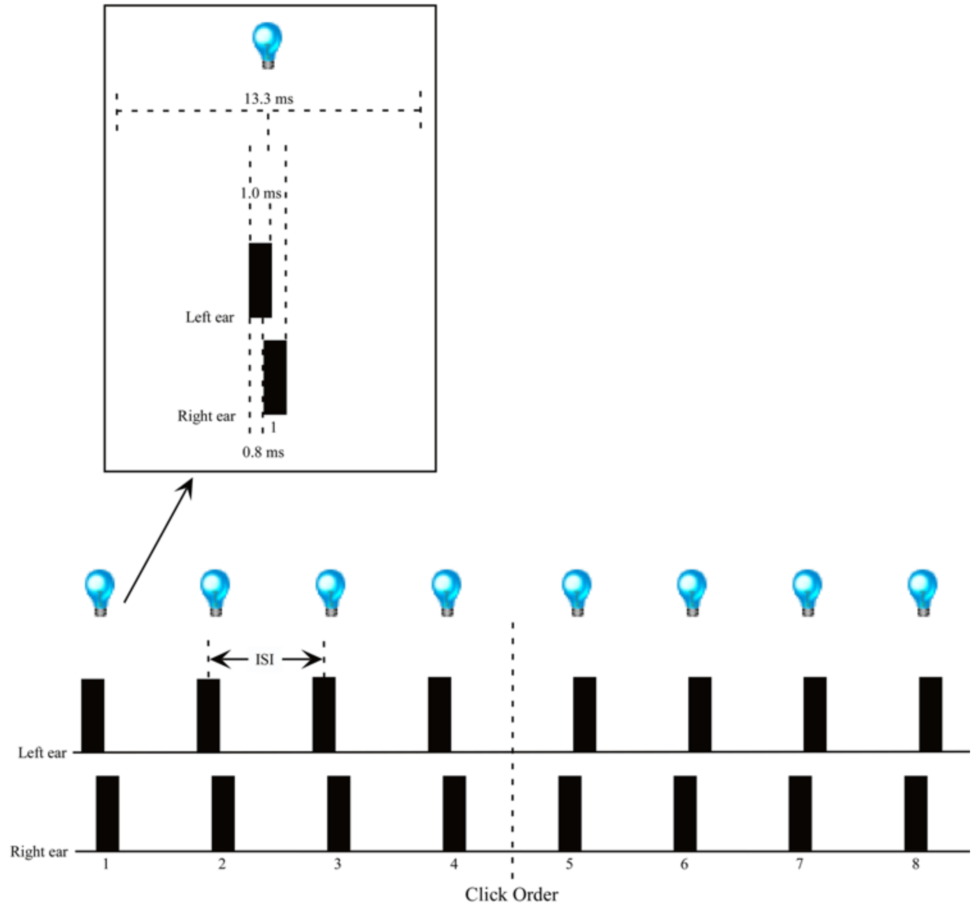

Fig. 2 Sequence of eight click-pairs. The first four are left-ear leading, while the second four are right-ear leading. The *light bulbs* represent the presentation of visual stimuli from Fig. 1 (not to scale)

Localization: exogenous (circles)

Localizations in the exogenous experiment did not differ significantly from those in the endogenous experiment ($F = 0.066$, $df = 79$, $p = .799$, $\eta^2 = .001$) (see Fig. 4). Moreover, the average range for left and right-ear-leading click-pairs in the non-conflicting condition was 10.9 units, in the control condition it was 8.4 units, and in the conflicting condition it was 6.0 units.

Localization: endogenous (numbers)

There were six conditions in total. These included two non-conflicting conditions: left-ear leading with the visual cue beginning at the left (left–left) and right-ear leading with the visual cue beginning at the right (right–right), two conflicting conditions: left-ear leading with the visual cue beginning at the right (left–right) and right-ear leading with the visual cue beginning at the left (right-left), and two control conditions with either right-ear leading or left-ear leading and an uninformative visual cue on the vertical meridian, which did not move. Perceived locations in the non-conflicting conditions followed the direction of the auditory target. However, instead of being perceived as four clicks from one side followed by four clicks from the other side, participants' localizations showed a relatively equal change perceived between each click. The average range for left and right-ear-leading click-pairs was 10.4 units.

When compared to the control conditions with uninformative numbers, the non-conflicting signals were significantly different in the right-leading and left-leading conditions ($F = 31.321$, $df = 1, 78$, $p = .001$, $\eta^2 = .286$). The average range for left and right-ear-leading click-pairs in the control condition was 9.6 units.

In conditions with conflicting signals, this pattern was reversed. The left-ear leading condition was perceived to travel from right-to-left, and the right-ear leading condition was perceived to travel from left-to-right. This effect was not as strong as 9.6 units ($t = 2.14$, $df = 78$, $p = .008$, $\eta^2 = .678$). The average range for left and right-ear-leading click-pairs in the conflicting condition was 6.2 units.

Response time

Median response times were calculated for all conditions in the endogenous and exogenous experiments and compared using an independent samples *t* tests. Response times include reaction time to the stimuli as well as the time

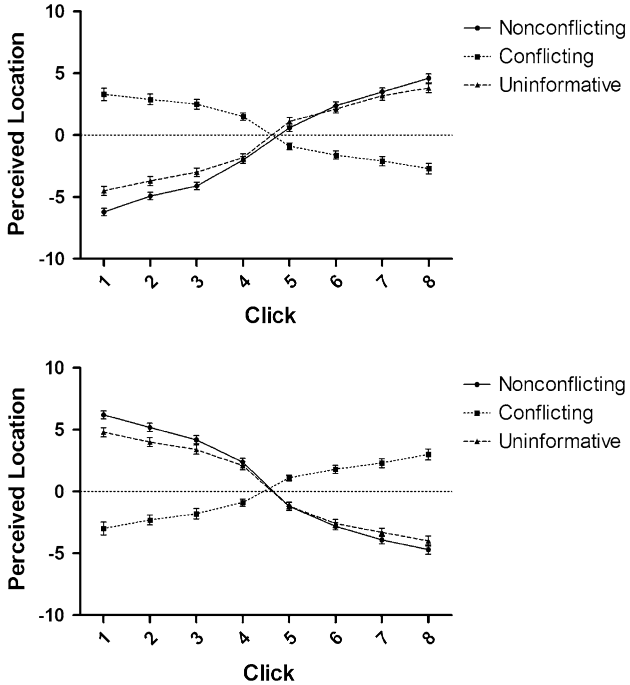

Fig. 3 Median perceived localizations of each click with endogenous non-conflicting visual and auditory stimuli (*solid line*), conflicting stimuli (*dottted line*), and an uninformative arrow (*dashed line*) in the left-ear leading condition (*top*) and the right-ear leading condition (*bottom*). *Error bars* equal SEM

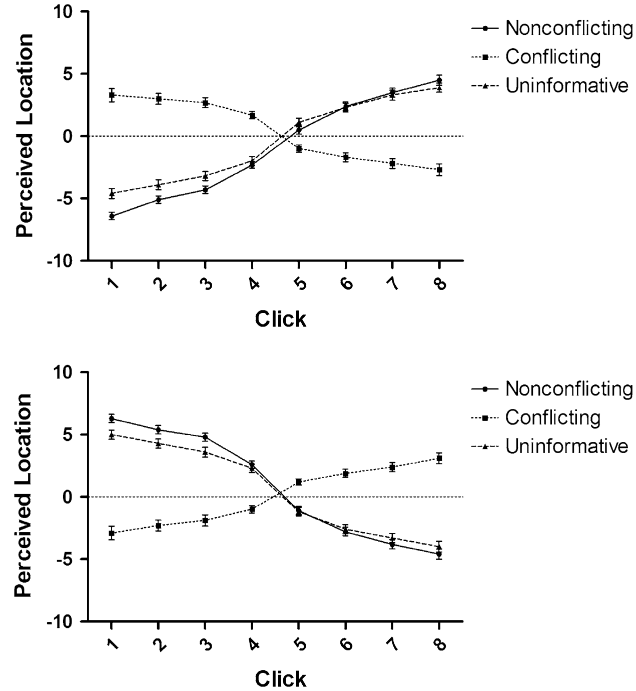

Fig. 4 Median perceived localizations of each click with exogenous non-conflicting visual and auditory stimuli (*solid line*), conflicting stimuli (*dotted line*), and an uninformative circle (*dashed line*) in the left-ear leading condition (*top*) and the right-ear leading condition (*bottom*). *Error bars* equal SEM

taken to make the sliding scale responses. While the latter condition is long, it is constant across conditions so that the only variable time in the response time is during the initial reaction time. Endogenous (circle) response times (median = 15.40 s) were shown to be significantly greater than exogenous (arrow) response times (median = 14.18 s) (t = 2.097, df = 78, p = .037, η^2 = .122). Moreover, endogenous (arrow) response times (median = 15.40 s) were shown to be comparable to exogenous (numbers) response times (median = 15.42 s) (t = 0.470, df = 78, p = .639, η^2 = .022).

Discussion

In essence, these findings support the notion that visual spatial cues bias auditory perception. However, this finding is unlike that found by Soto-Faraco et al. (2004) who discovered that visual modulation of auditory motion is caused by an illusory reversal in the perceived direction of sounds, in that what is novel in the current study is that the sounds are internal to the head. That is, the sound is perceived to step through the head while the visual signal is generated outside the head. Instead, consider the results that would be expected if the visual cue completely dominated perceptions of the auditory signal location (the strong version of visual capture theory). In this case, the "non-conflicting" and "conflicting" plots in the figures should together form perfectly symmetrical Xs. Yet, they do not. Instead, the "conflicting" case is a much flatter function than the "non-conflicting" case. Moreover, the "non-informative" data should all fall on a flat line. The fact that the latter effect does not occur demonstrates that there was a significant influence of the auditory signal.

The fact that an approximate X-like pattern was observed, and especially that the visual cue reverses the perceived spatial location order in the "conflicting" condition, tends to support visual capture theory. However, visual capture is far from complete. There is also some influence of the auditory signal. Therefore, a theory based on cue combination or compromise between localization based on visual and auditory inputs probably provides the best account of the results, even if the two types of inputs do not contribute equally to auditory localization, as the supramodal theory suggests. The fact that the data from the non-conflicting conditions conform to smoothed step function is also consistent with a weighted combination of evidence from visual and auditory inputs, as is the fact that the range of perceived locations is constricted in the conflicting conditions relative to the non-conflicting conditions.

Second, an examination of the response time data further supports the theory of visual capture. Response times for endogenous conditions (i.e., arrows and numbers) were significantly greater than response times for exogenous conditions (i.e., circles) as was found in the related work of Chakravarthi and VanRullen (2011). This is due to the time required to transfer one's attentional window to the location of the tips of the arrows or to the numbers in the endogenous conditions. Since the localization data were comparable for endogenous and exogenous conditions but the response time data were different, these experiments support visual capture. The factors influencing the relative weights given to each type of cue are still unclear. We suspect that any conclusions reached on the basis of this data set might depend rather heavily on parametric aspects of the presentation, such as the relative timing of the visual and auditory cues. A deeper analysis would take into consideration processing times for visual and auditory information, some of which may be known from prior work on short-term memories. We would like to consider pursuing this direction in future work.

It is clear that virtually all studies that favor visual capture have a more salient visual signal; in fact, it is almost impossible to degrade the visual signal enough to alter this bias (Hairston et al. 2003). Thus, our visual signal is probably more salient, and this is what may drive visual capture. Additionally, these experiments demonstrated illusory directional hearing in all conditions (Hari 1995). This finding contributes to our understanding of this phenomenon by showing that illusory directional hearing can be produced in the presence of both conflicting and non-conflicting visual stimuli. Through the use of an illusory directional hearing paradigm (as opposed to one in which participants were presented with external sounds), we were able to demonstrate that visual stimuli in external space can influence internal measures of localizations within the head. This finding suggests that visual capture is a generalizable and robust phenomenon.

In the present study, eye movements were not measured. However, it is believed that eye movements should not be a confounding factor. The arrows were located in the center of the screen; therefore, only their rotation matters which does not require eye movement, yet it results in a shift in attention. Additionally, the presentation of visual and auditory stimuli in this study was set at a rate (100 ms) faster than what would be required for a typical eye movement to occur (about 200 ms).

Acknowledgments Special thanks to John Petrocelli for help programming the experiments.

References

Chakravarthi R, VanRullen R (2011) Bullet trains and steam engines: exogenous attention zips but endogenous attention chugs along. J Vision 11(4):121–122. doi:10.1167/11.4.12

Farah MJ, Wong AB, Monheit MA, Morrow LA (1989) Parietal lobe mechanisms of spatial attention: modality-specific or supramodal? Neuropsychologia 27(4):461–470. doi:10.1016/0028-3932(89)90051-1

Hairston WD, Laurienti PJ, Mishra G, Burdette JH, Wallace MT (2003) Multisensory enhancement of localization under conditions of induced myopia. Exp Brain Res 152(3):404–408. doi:10.1007/s00221-003-1646-7

Hari R (1995) Illusory directional hearing in humans. Neurosci Lett 189:29–30. doi:10.1016/0304-3940(95)11443-Z

Hay JC, Pick HL, Ikeda K (1965) Visual capture produced by prism spectacles. Psychon Sci 2(8):215–216

Mays A, Schirillo J (2005) Lights can reverse illusory directional hearing. Neurosci Lett 384:336–338. doi:10.1016/j.neulet.2005.05.021

Soto-Faraco S, Spence C, Kingstone A (2004) Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities. J Exp Psychol Human 30(2):330–345. doi:10.1037/0096-1523.30.2.330

Spence C, Driver J (1996) Audiovisual links in endogenous covert spatial attention. J Exp Psychol Human 22(4):1005–1030. doi:10.1037/0096-1523.22.4.1005

Ward LM (1994) Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. Can J Exp Psychol 48(2):242–259. doi:10.1037/1196-1961.48.2.242