

Visual-Spatial Attention: Can Sounds Affect Perception of Visual Objects?

Charlotte Clark

University of Essex

Abstract

It is currently understood that it is possible for attention to be drawn by a stimulus to an area in space without the conscious decision to do so and that this can be achieved across modalities; for example, a vibration of a mobile phone or a ringtone can draw eyes to the phone's location (Hairston et al., 2003). Whether a combination of audio and visual information can affect perception of visual objects enough to demonstrate inhibition of return (IOR) is unknown. In a classic Posner task (1980), participants fixated at the centre of a display screen and had their attention drawn to surrounding areas. When a flash is cued at a surrounding location, target detection was facilitated comparative to anywhere. In a stream/bounce paradigm (Sanabria et al., 2004), motion and sound was added to a similar kind of display to that of Posner's (but without cueing), to investigate the perceptual effects of combined visual and auditory information on attention. It was discovered that the addition of a sound could influence how the display was perceived. With a combination of the Posner (1980) display and stream bounce paradigm, it is possible to test whether the perceptual influences of sound can change the dynamics of how attention is known to work. In this experiment, object-based IOR was not found when just visual stimuli were presented, but was evident when visual and auditory information were combined. The results demonstrate that there is promise for further research into the influence of sound on visual perception, but that stimulus driven spatial attention can be sensitive to

methodological issues arising from the specifics of the task. Consequently, spatial interference and eye movements are discussed.

Keywords: Visual-Spatial Attention, Eye Movements, Spatial Interference, Attention

Introduction

The ability to pay visual attention in an environment and combine information with other incoming senses (e.g. auditory information) comes as a multisensory, evolutionary advantage. This is because localising a snake, for example, is easier when combining spatial information (movement in the undergrowth with a 'rattle' sound), than locating a predator with a single modality on its own. Attentional mechanisms therefore are adapted to be integrated so that they can pick out and combine the saliency of any given environment as needed even without the conscious control of the observer. In the visual domain, space-based attention, as demonstrated by Posner (1980), is an attentional prioritisation of a location. Posner's well known cueing paradigm used a central fixation point and peripherally located objects to successfully evidence that when fixating in the centre of a display screen, exogenous cueing (any non-informative deliberate event before the target onset that captures attention) can manipulate how quickly a spatial location can be attended to without direct eye movements. Thus, it is possible to attend psychologically to an area (using covert attention) without direct eye contact. Subsequent investigation demonstrated the variability of spatial attention. It was likened to a metaphorical spotlight, able to adapt and accommodate attentional spread according to the task (Eriksen and Eriksen, 1974). Thus, local features can either help or hinder depending on their similarity, or their salience (LaBerge, 1983). Furthermore, Egly, Driver and Rafal (1994) demonstrated that spatial attention was object based since reaction times were quicker for targets within the same object than an alternative object. Therefore, attention spreads within the cued object fastest. However, beyond a 250 millisecond (msec) timeframe, the benefit is reversed and response delayed, a concept known as inhibition of return (Klein and Ivanoff, 2008). During an IOR task, participants fixate on an uninformative central location within a visual display, and reaction time to locate a target appearing peripherally is measured. Sometimes the target is correctly cued and at other times, not. Consequently, it is possible to manipulate which side will be responded to quickest because IOR enables new stimuli to be prioritised over old by

the use of inhibitory tagging (Klein and Ivanoff, 2008). Further, moving objects equally show this tendency (Tipper, Driver and Weaver, 1991). Since object movement does not affect this preference, this study will consider what happens to IOR when the object's identity and location are ambiguous. Moreover, with this ambiguity, can spatial attention be manipulated cross-modally by the addition of sound? For instance, visual motion perception is altered by the addition of sound when discs approach one another from either end of a visual display and the point of coincidence is occluded. When no additional stimulus is present, the discs are perceived as passing (streaming). However, a sound input at the moment of coincidence changes perception to discs bouncing off each other (Sekuler, Sekuler and Lau, 1997). This is because the senses are designed to cooperate (Spence et al., 2000). Interestingly, IOR has equally been demonstrated as supramodal in this way (Spence et al., 2000). Thus, since IOR can travel with an object (Tipper et al., 1991) and motion perception can be altered by sound, it leaves open the question of whether IOR could be manipulated to travel with an object during a bounce/stream paradigm. This investigation will seek to identify whether a combination of sound and vision can affect how objects are perceived and attended to. To achieve this, a typical Posner cueing task (1980) will be adapted into the bounce/stream paradigm demonstrated by Sanabria, Correa, Lupianez and Spence (2004). If IOR tends to follow this stream/bounce perception, it is expected that inhibition should apply to different sides depending on whether or not the sound is present. With the sound, inhibition should return to the starting location and RTs should be slower when cue and target are on the same side of the screen.

Method

Participants

An opportunity sample of sixty-eight healthy volunteers, 17–50 years of age (mean age 21 years), took part in the study. All were naive to the purposes of the experiment. Six participants were excluded (see results section). The remaining sixty-two participants consisted of fifty-seven right handed and five left handed individuals. All reported having normal hearing and normal or corrected-to-normal vision.

Apparatus

The Stimulus presentation, response times (RT) and error rates were controlled and recorded by a Mac computer (Superlab) and keyboard. Headphones delivered the sound.

Materials

The display layout can be seen in *Figure 1*. Display measurements were as follows: Discs 1.2 cm (diameter), distance of discs to top and bottom of display 6 cm, from discs to occluder 8 cm, visual occluder 9.7 cm x 4.7 cm, distance of occluder to top and bottom of display 6 cm, display box 23.5 cm x 13.5 cm, central fixation 1.1 cm x 1.1 cm.

Design

A 2x2 within subjects design was used. The first factor was block type with two levels: sound or silent. The second factor was side of target with two levels: same (screen side as cue) or different (opposite screen side to cue) randomised throughout a block. The dependent variable was RT in msec. The order of the blocks was decided beforehand (by coin toss) and was alternated between sound and silent conditions for a total of 4 blocks.

Procedure

Care was taken to ensure that participants did not see the experiment beforehand and participants were advised that the experiment was designed to see whether what they hear and see affected how quickly they reacted to a visual target. All participants were tested in individual booths and given the same information. Participants chose a preferred responding hand in advance and were told to react as quickly and accurately as possible. Clear instruction was given to focus on the fixation point (central black cross) throughout trials and to have their chosen hand readied at the keyboard. For each participant, a ten trial practice was given to familiarise themselves with the task, in same block-type as had been allocated as the starting block. If more practice was required, more trials were given. Participants wore headphones in all blocks.

For silent trials, a white horizontal rectangle was centrally placed within a black screen. Inside were black discs either side of the screen equidistant from the centre (as shown in *figure 1*). At the start

of every trial the display was shown for 500 msec before a non-informative white circle (cue) flashed for 200 msec within one of either discs to initiate covert attention. Afterwards, the discs moved to the opposite side of the screen taking 1000 msec (*figure 1, T3*). Whilst occluded, (*figure 1, T2*) the fixation momentarily flashed. On same side trials, the target (white asterisk) subsequently appeared on the same side spatially within a disc. In the different side trials, the target appeared on the opposite side spatially. On catch trials no target appeared and after 2 seconds (if correctly rejected), the trial terminated. For half of the trials the target appeared immediately after movement stopped, whilst for the other half there was a 200 msec delay. Participants responded with the keyboard space bar ending the trial and initiating the next.

For the sound block the same format was used, but during occlusion the sound (click) was delivered for 100 msec at the coincidence point of the discs. Each block contained a mixture of same, different and catch trials randomly presented. Each block had sixty trials, twenty-four same/different side trials randomised amongst twelve catch trials.

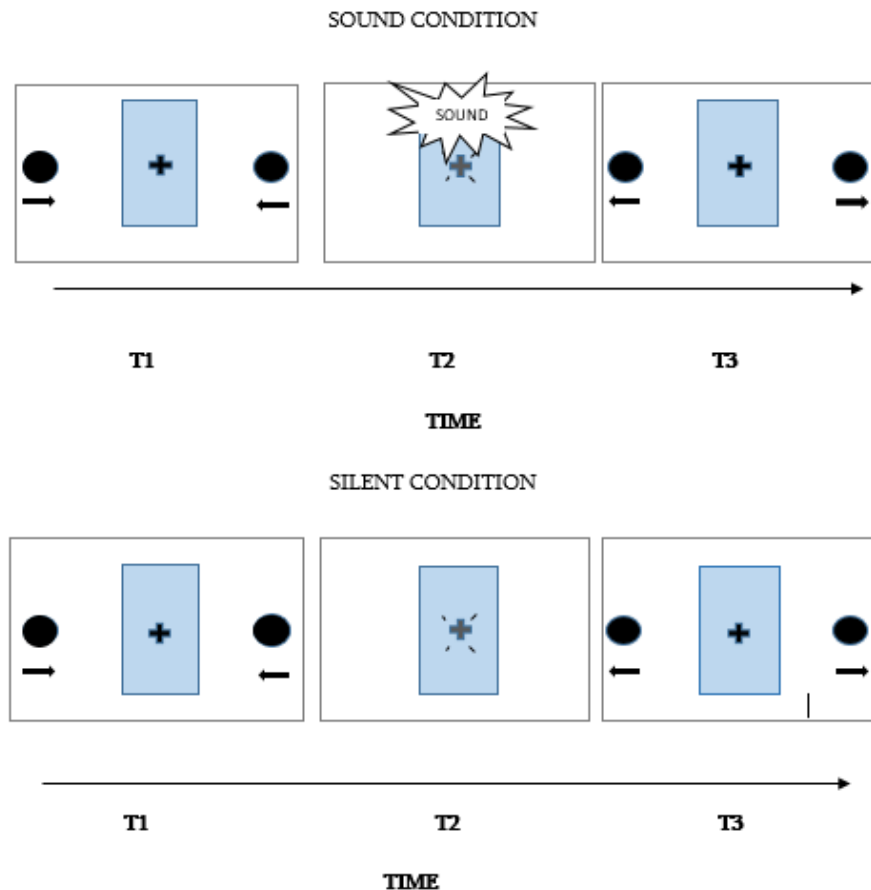


Figure 1 A schematic illustration for a sound trial (encouraging the perception of a bounce), and the silent trial (encouraging the perception of streaming). The black arrows below the discs indicate the direction of movement (before and after the occlusion) for that trial. T1, onset of motion at the beginning of the trial (after the cue has been deployed) T2, the occlusion point with central fixation point flashing, T3, the re-emergence of discs after coincidence.

Results

The data from six participants was excluded from analysis; four were omitted due to a lack of understanding and two were omitted due to a failure to complete. Trials where the target was not present (catch trials) were placed into four categories and counted. The trials were grouped according to whether or not the sound was present, those that were accurately ignored (correct rejections) and those that were accidentally responded to (false alarms). The number of false alarms was 4.1%. The trials where the target was present were counted using the criterion that the response was made between 50–1500 msec after the target onset. These trials were counted (as hits) and the rest as misses (2.6% misses overall). Target trials were grouped according to whether the

target and cue had appeared on the same, or on separate sides. For each participant and each combination of sound (present/absent) and side cue (same/different), the median reaction time (RT) was calculated using only the trials that were hits. Overall, four scores were calculated for each participant. The inter-participant means of median RTs were then calculated for each condition (see figure 2).

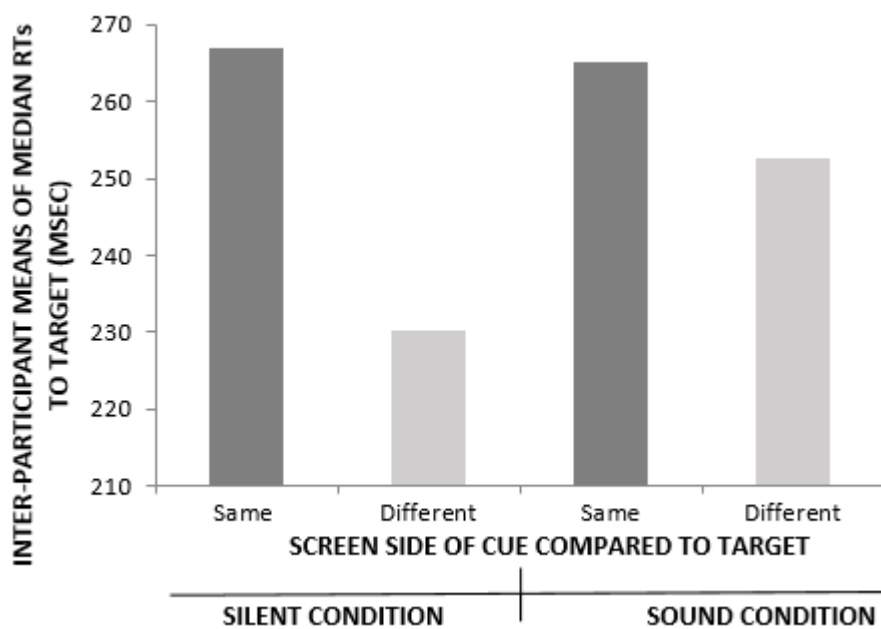


Figure 2 Inter-participant means of median reaction times to locate a same screen side (dark grey) versus different screen side (light grey) cued target by condition (sound versus silent). Silent condition standard deviations were 68.5 (same), 65.7 (different).

A two-way (*sound* by *side*) within subjects ANOVA was conducted with mean of median RT as the dependent variable. There was a significant main effect of condition (*sound* vs *silent*) $F(1, 61) = 6.65$ $p = .013$, and a significant main effect of side (*same* vs *different*) $F(1, 61) = 134.98$ $p < .001$. There was also a significant interaction of *sound* by *side*, $F(1, 61) = 51.13$ $p < .001$. Pair-wise planned comparisons (*t*- tests) revealed that participants' RTs were significantly faster for *different* side trials compared to *same* side trials in the *silent* condition, $t(61) = 12.30$ $p < .001$, showing no object-based IOR in the *silent* condition as responses were quickest for the cued object (in new location). Observers were also significantly faster to react to *different* side trials compared to *same* side trials in the *sound* condition $t(61) = 5.24$ $p < .001$. Successfully evidencing IOR for the object after a perceptual bounce facilitated by sound. Thus, sound may be able to influence IOR in visual

perception, but without concise evidence of object-based IOR in the *silent* condition, this cannot be inferred.

Discussion

The aim of the study was to investigate whether the addition of sound could alter visual perception. To achieve this, the Posner style cuing task (1980) was modified with a bounce/stream paradigm Sanabria et al. (2004), in an attempt to influence the perception of IOR. Contrary to expectation, the results yielded no evidence of IOR (consistent with a stream) as evidenced by Sanabria et al. (2004), since object based IOR should have travelled with the object creating an advantage for the side originally cued. This would be consistent with object-based IOR as evidenced by Tipper et al. (1991). Rather, observers were faster on the different side (with the cued object). Interestingly, for the sound condition, IOR was demonstrated (in line with the perception of a bounce) suggesting that IOR can be influenced by the addition of sound. Unfortunately, without coherence between conditions this cannot be concluded. Therefore, further consideration of the results is required.

Given that IOR is a known component of spatial attention that can be produced in most attentional experiments (Klein and Ivanoff, 2008) and that it still applies even when objects move (Tipper et al., 1991) and can be shown between different sensory modalities, (Spence et al., 2000), it is likely that methodological issues account for the findings. One contingency could be eye position. The central fixation point is an essential element in all covert attention tasks and whilst the importance of central fixation was conveyed, eye movements were not tracked and therefore cannot be relied upon. Posner and Cohen (1984) for example demonstrated that IOR requires that attention is drawn back to the fixation point after cuing. Consequently, if eyes move within the trial, the inhibitory effect could remain at the environmental position of where the cue had occurred. This is consistent with both findings, and consistent with the reduced effect (*figure 2*) evidenced in the sound compared to the silent condition since, regardless of the perceptual set, the same side (cue to target) would be inhibited if this was the last place of energy. Thus in the sound condition, the findings, may not have arisen from multisensory visual perception, but from failure to remain fixated.

Equally, Tipper et al. (1991) evidenced that IOR in static displays has an additional inhibitory component compared to moving displays. Thus with moving displays, covert attention may be more easily disrupted since object-based attention requires additional mental resources (Chen, 2012). For example, during Tipper et al. (1991) moving paradigm, the objects rotated around the fixation point remaining at an equidistant point from fixation at all times. Moreover, at no point were they occluded. In the current paradigm however, the objects were occluded behind the fixation area. Therefore, if the inhibitory tagging was disrupted during the period of occlusion IOR may have diminished. This is possible because ambiguity of movement has been highlighted as a potential pitfall in moving displays (Reppa, Schmidt and Leek, 2012). Tipper, Brehaut and Driver (1990) however, evidenced object based IOR with occluding columns. Therefore, interference may have increased in this instance, compared to Tipper et al. (1991) by the proximity of the objects to fixation, which changed throughout the task substantially and even connected with the object in which fixation was presented.

Importantly, the substantial changes in movement could have altered the attentional spread beyond fixation to include the incoming discs (Eriksen and Eriksen, 1974), especially since the similarity of surrounding features is known to interfere (LaBerge, 1983) and that distance between objects is also crucial (Franconeri, Johnathan and Scimeca, 2010). Thus, without any extra perceptual cues available, and coupled with objects similarity, it is conceivable that these factors compromised covert attention more markedly in the silent block compared to the sound block. Moreover, the task combination may have diminished the inhibitory tagging (Egly, et al., 1994) by interference (Franconeri et al., 2010). However, without tracking eye movements, this claim cannot be substantiated.

Evidencing eye movements was one of the criticisms of Spence et al. (2000) during his supra-modal paradigm; the use of eye tracking with sub-set of participants in a further replication of this study would be beneficial. More importantly, with eye movements accounted for, the other methodological issues such as ambiguity of movement across fixation could be considered. However, as demonstrated in other cross-modal investigations (Kennett, Spence and Driver, 2002), visual restriction can be an important feature in studying multisensory perception. Indeed, for the perception of a stream/bounce effect, the visual restriction is warranted. Thus, it may be possible to

just relocate the ambiguity to a location away from fixation. The movement would still need to be directed across both left and right visual hemifields (Sereno and Kosslyn, 1991), so as not to increase the difficulty of the task. However, the space between objects and fixation could be changed to reduce the likelihood of interference (Franconeri et al., 2010). For example, Tipper, Jordan and Weaver (1999) evidenced scene-based inhibition at different multiple locations on the screen; thus, if the display was moved equidistantly above fixation, it may be possible to eradicate or at least re-evaluate the aforementioned possible limitations.

Conclusion

In conclusion, the demonstration of IOR in a cross modal attention paradigm such as this can be difficult to produce. However, the results show promise towards the cross-modal influence of sound upon spatial attention. Given that the display used has important methodological constraints, a change in the spatial qualities of the display is needed to rule out interference between fixation and covert attention. Equally, since eye movements were not ruled out, future investigations would need to also factor eye tracking into the methodology. With these alterations, and the eventual observation of the object-based IOR, it would be possible to advance the understanding on how the influence of sound can affect perception and attention. In addition, it would facilitate better understanding of the cross modal elements of spatial attention more generally.

References

- Chen, Z. (2012) 'Object-based attention: A tutorial review', *Attention Perception & Psychophysics*, 74(5), pp. 784-802.
- Duncan, J. (1984) 'Selective attention and the organization of visual information', *Journal of Experimental Psychology: General*, 113(4), pp. 501-517.
- Egley, R., Driver, J. and Rafal, R. D. (1994) 'Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects', *Journal of Experimental Psychology: General*, 123(2), pp. 161-177.

Visual-Spatial Attention: Can Sounds Affect Perception of Visual Objects?

Eriksen, B. A. and Eriksen, C. W. (1974) 'Effects of noise letters upon the identification of a target letter in a nonsearch task', *Perception & Psychophysics*, 16(1), pp. 143-149.

Franconeri, S. I., Johnathon, S. V. and Scimeca, J. M. (2010) 'Tracking multiple objects is limited only by object spacing, not by speed, time or capacity' *Psychological Science*, 21(7), pp. 920-925.

Hairston W. D. et al. (2003) 'Visual localisation ability influences cross-modal bias', *Journal of Cognitive Neuroscience*, 15(1), pp. 20-29.

Kennett, S., Spence, C., and Driver, J. (2002) 'Visuo-tactile links in covert exogenous spatial attention remap across changes in unseen hand posture', *Perception & Psychophysics*, 64(7), pp. 1083-1094.

Klein, R. M. and Ivanoff, K. (2008) 'Inhibition of return', *Trends in Cognitive Sciences*, 4(4), pp. 138-147.

LaBerge, D. (1983) 'The Spatial Extent of Attention to Letters and Words', *Journal of Experimental Psychology-Human Perception and Performance*, 9(3), pp. 371-379.

Posner, M. I. (1980) 'Orienting of attention', *Quarterly Journal of Experimental Psychology*, 32(1), pp. 3-25.

Posner, M. I., & Cohen, Y. (1984) 'Components of visual orienting', in: Bouma, H. and Bouwhuis, D. (eds.) *Attention and Performance X*. London: Lawrence Erlbaum Associates, pp. 531-556.

Reppa, I., Schmidt, W. C., and Leek, E. C. (2012) 'Successes and failures in producing attentional object-based cueing effects', *Attention Perception & Psychophysics*, 74(1), pp. 43-69.

Sanabria, D. Et al. (2004) 'Bouncing or streaming? Exploring the influence of auditory cues on the interpretation of ambiguous visual motion', *Experimental Brain Research*, 157(4), pp. 537-541.

Sekuler, R., Sekuler, A. B. and Lau, R. (1997) 'Sound alters visual motion perception', *Nature*, 385(6614), p. 308.

Sereno, A. B. and Kosslyn, S. M. (1991) 'Discrimination within and between hemifields: a new constraint on theories of attention', *Neuropsychologia*, 29(7), pp. 659-675.

Visual-Spatial Attention: Can Sounds Affect Perception of Visual Objects?

Spence, C. et al. (2000) 'Inhibition of return is supramodal: A demonstration between all possible pairings of vision, touch and audition', *Experimental Brain Research*, 134(1), pp. 42-48.

Tipper, S. P., Brehaut, J. C. and Driver, J. (1990) 'Selection of moving and static objects for the control of spatially directed action', *Journal of Experimental Psychology: Human Perception and Performance*, 16(3), pp. 492-504.

Tipper, S. P., Driver, J. and Weaver, B. (1991) 'Object-centred inhibition of return of visual attention', *Quarterly Journal of Experimental Psychology A*, 43(2), pp. 289-298.

Tipper, S. P., Jordan, H. and Weaver, B. (1999) 'Scene-based and object-centred inhibition of return: Evidence for dual orienting mechanisms', *Perception & Psychophysics*, 61(1), pp. 50-60.