Evidence for Crossmodal Interactions across Depth on Target Localisation Performance in a Spatial Array

Article *in* Perception · November 2012

DOI: 10.1068/p7230

CITATIONS
7

READS
143

5 authors, including:

Jason Seeho Chan
University College Cork
44 PUBLICATIONS **430** CITATIONS
SEE PROFILE

Corrina Maguinness
Technische Universität Dresden
24 PUBLICATIONS **300** CITATIONS
SEE PROFILE

Annalisa Setti
University College Cork
119 PUBLICATIONS **1,200** CITATIONS
SEE PROFILE

Fiona N Newell
Trinity College Dublin
159 PUBLICATIONS **4,812** CITATIONS
SEE PROFILE

Perception, 2012, volume 41, pages 757 – 773

doi:10.1068/p7230

Evidence for crossmodal interactions across depth on target localisation performance in a spatial array

Jason S Chan[1,2], Corrina Maguinness[1], Danuta Lisiecka[1], Annalisa Setti[1], Fiona N Newell[1]

[1] School of Psychology and Institute of Neuroscience, Trinity College Dublin, Ireland; [2] Institute for Medical Psychology, Goethe University, Heinrich-Hoffmann Strasse 10, Frankfurt am Main, Germany; e-mail: chan@med.uni-frankfurt.de

Received 8 February 2012, in revised form 12 June 2012

Abstract. Auditory stimuli are known to improve visual target recognition and detection when both are presented in the same spatial location. However, most studies have focused on crossmodal spatial congruency along the horizontal plane and the effects of audio-visual spatial congruency in depth (ie along the depth axis) are relatively less well understood. In the following experiments we presented a visual (face) or auditory (voice) target stimulus in a location on a spatial array which was either spatially congruent or incongruent in depth (ie positioned directly in front or behind) with a crossmodal stimulus. The participant's task was to determine whether a visual (experiments 1 and 3) or auditory (experiment 2) target was located in the foreground or background of this array. We found that both visual and auditory targets were less accurately located when crossmodal stimuli were presented from different, compared to congruent, locations in depth. Moreover, this effect was particularly found for visual targets located in the periphery, although spatial incongruency affected the location of auditory targets across both locations. The relative distance of the array to the observer did not seem to modulate this congruency effect (experiment 3). Our results add to the growing evidence for multisensory influences on search performance and extend these findings to the localisation of targets in the depth plane.

Keywords: multisensory perception, distance perception, audiovisual spatial perception

1 Introduction

Previous research has provided evidence for important cross-sensory interactions in the localisation of a unisensory target. For example, it is well known that the visual system can provide a spatial frame-of-reference which can facilitate the localisation of an auditory target along the azimuth (Alais and Burr 2004; Jackson 1953; Shelton and Searle 1980; Warren 1970). Similarly, an auditory stimulus can aid the detection of a visual stimulus when both are presented from the same spatial location (Perrott 1984; Perrott et al 1995; Spence 2007; Spence and Driver 2000). In particular, these studies showed an effect of spatial congruency by manipulating the relative locations of audio-visual stimuli along the horizontal plane only. For example, Perrott and colleagues asked participants to search for a visual target while presenting a spatially congruent or incongruent auditory stimulus (relative to the visual target) or no sound. They found that a spatially congruent auditory stimulus improved visual target detection, compared to the incongruent sound or sound absent conditions. However, in their study, and other subsequent investigations of crossmodal influences on target localisation (eg Spence and Driver 2000), regardless of the congruency condition, the auditory stimulus was always presented at the same distance from the observer as the visual target and inter-stimulus distance was manipulated along the horizontal plane only. Since these studies did not investigate any possible effects associated with disparate audio-visual stimuli along the depth plane (ie along the depth axis), it remains unclear whether crossmodal stimuli along this spatial dimension would also affect target localisation in either modality.

Previous studies suggest there may be important interactions between the modalities when perceiving targets positioned in depth. For example, in an investigation of distance perception, Loomis et al (1998) found visual depth perception to be significantly better than auditory depth perception when participants were asked to walk to a target positioned directly in front of them. In their study, participants either saw a target or heard a sound emitted from a target loudspeaker (while wearing a blindfold) at various distances, and their task was to walk to the location of the target. This difference in performance across the modalities may be attributed to the relatively richer information available to the visual system regarding the spatial layout, environment, and relative distance cues, all of which were not easily available in the auditory domain. As a confirmation of this idea, Philbeck and Loomis (1997) found that accuracy in perceiving visual depth decreases as the number of visual cues are reduced. Furthermore, Zahorik (2001) reported that when visual information regarding the spatial layout of a room and testing apparatus was provided, the perceived distance of an auditory target was more accurate relative to when no visual information was presented. Specifically, Zahorik reported that blindfolded participants underestimated the distance of auditory targets more often than those participants who could see the first loudspeaker as a visual reference. Therefore, the perceived distance of an auditory stimulus appears to be affected by the location of a visual reference stimulus. Indeed, the visual capture of the location of an auditory stimulus is best known either as the 'ventriloquist illusion' (Howard and Templeton 1966) or the 'proximity image effect', if visual capture occurs in depth (Gardner 1968)—although, again, most investigations of this illusion have typically limited the distances between the auditory and visual stimuli to the horizontal rather than the depth plane (see Bertelson and de Gelder 2003 for a review; Bertelson et al 2000). Interactions between vision and sound along the depth axis are important to establish as it is well known that distance can be underestimated in the visual (Fukusima et al 1997; Loomis et al 1992; Plumert et al 2004; Ziemer et al 2009) and auditory modalities (Kearney et al 2010; Loomis et al 1998), with auditory distance consistently more underestimated than visual distances (Loomis et al 1998; Zahorik 1998, 2001).

While there are several studies which have explored the spatial interactions between the senses, others have also shown that temporal factors can also play a role on cross-modal spatial perception (Alais and Carlile 2005; McDonald et al 2000; Sekuler et al 1997; Shams et al 2000; Sugita and Suzuki 2003). For example, Van der Burg et al (2008) reported that in a dynamic search display, with visual target and distractors randomly appearing and disappearing, a sound which is synchronised with the onset of the visual target can facilitate its detection. They argued that the temporal coincidence between sound and vision allowed for the integration of these features which, in turn, rendered the target more salient in the display. However, the display was presented in the picture plane; therefore, it is unknown whether sounds presented in depth would have the same effect on visuo-spatial localisation. Indeed, Sugita and Suzuki (2003) reported that distance from the observer can affect the perceived temporal order between auditory and visual stimuli (temporal order judgment or TOJ). In their study, visual stimuli were presented at various distances and the auditory stimuli, which were filtered with head-related transform functions to simulate distance in the auditory domain, were presented through headphones. They found that the temporal window between vision and audition was shifted by 3 ms m^{-1}—that is, it followed the speed of sound (Sugita and Suzuki 2003). Sugita and Suzuki (2003) suggest that the temporal window of integration between vision and audition is elastic and can be modulated by increasing distance from the observer (see also Alais and Carlile 2005 for similar findings). The perceived spatial location of the sound relative to the visual stimulus was not explicitly tested, so it is difficult to determine the extent to which the perceived relative or 'absolute' auditory distance perception affected the temporal window of integration.

In fact, some researchers have suggested that head-related transform functions provide little useful distance information (Zahorik 2002). In any case, the results of Sugita and Suzuki provide evidence that the perceived relative distance of sounds can affect the perceived temporal order of crossmodal simulation across various distances, suggesting that spatial and temporal effects interact in the perception of multisensory spatial events.

The current study was designed to extend the previous studies on audio-visual interactions in target localisation by investigating the role of location along the azimuth and distances in the depth plane between auditory and visual events. Specifically, using a relatively complex 2-dimensional display, we explored the role of spatial depth information when localising a target in one modality when information from another modality was also simultaneously presented. Across three experiments, the task for the participants was to identify the spatial location of either a unimodal (visual or auditory) target in a spatial array. We hypothesised that, if spatially congruent information is necessary for efficient multisensory perception, then spatially incongruent audio-visual stimuli along the depth plane would impair performance in localising a target. Furthermore, given that vision is considered to be the modality with the highest spatial precision (Welch et al 1986; Welch and Warren 1986), particularly for centrally presented targets, and that vision is known to capture the location of an auditory stimulus (Gardner 1968; Howard and Templeton 1966), it was expected that spatially incongruent distractor sounds would have less of an effect on disrupting localisation performance in the centre of the array than in the periphery. In contrast, we predicted that performance in localising auditory targets would be more disrupted by spatially incongruent visual distractors when both were presented in central positions than in peripheral positions in the array.

2 Experiment 1

The following experiment investigated localisation performance to spatially congruent and incongruent audio-visual stimuli in depth. Visual targets consisted of an image of a face, whereas auditory stimuli consisted of a voice saying 'Hi'. In an array of 8 possible locations of visual targets, a single target was indicated by a light which backward lit the face stimulus. An auditory stimulus was presented either at the same location as the visual target (congruent condition) or at a different location in depth, either in front of or behind the visual target (incongruent condition), or no sound was presented. We expected to find a similar interference effect of auditory distractor stimuli, which were incongruently located in depth on visual target localisation, as was previously found in studies where auditory and visual targets were presented along the horizontal plane (eg Perrott et al 1995).

2.1 *Method*

2.1.1 *Participants.* Twenty-nine (nineteen female) participants between the ages of 17 and 36 years (their mean age was 24 years) took part in this experiment. All, bar one, of the participants reported being right-hand dominant, and all participants reported normal or corrected to normal vision and no hearing impairments. This experiment (and all subsequent experiments) was approved by Trinity College School of Psychology Research Ethics Committee, and accordingly all participants provided informed written consent prior to testing.

2.1.2 *Apparatus and materials*. The apparatus consisted of an array of eight loudspeakers (FRWS 5: Visaton, Germany), each with a diameter of 5 cm and mounted onto separate wooden poles at a height of 120 cm. These poles with loudspeakers were positioned on the floor of a testing laboratory, arranged in two semi-circular arrays. The foreground array was positioned at a distance of 60 cm and the background array at 120 cm away from the

participant, who was seated at the centre of the array. Each stimulus was positioned 45° away from its neighbouring stimulus on the horizontal array. As such there were two peripheral locations of the visual/auditory targets: left and right 67.5° from fixation, and two more central locations left and right of fixation by 22.5° (see figure 1 for an illustration of the apparatus). The auditory stimulus was a recording of a male voice saying 'Hi'. The sound pressure level of each loudspeaker was equated at source and relative to the participant, across the foreground and background array. In other words, for each loudspeaker in the 60 cm (foreground) array the sound pressure level was approximately 68 dbA, and for each loudspeaker positioned in the 120 cm (background) array the sound pressure level was 65 dbA. Faces and voices were chosen as stimuli because faces and voices are familiar bi-modal stimuli. Moreover, the familiarity of an auditory stimulus is considered to be an important factor for 'absolute' auditory depth perception (Coleman 1963; Loomis et al 1998).

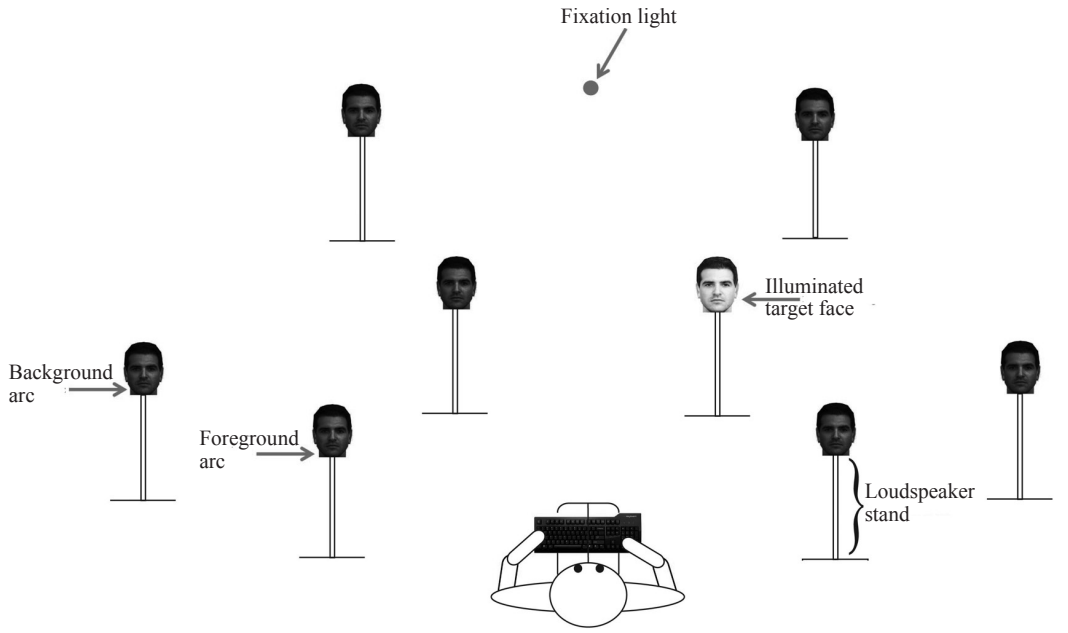

Figure 1. An illustration of the target array. Each stimulus was presented in either the foreground or background array relative to the participant. Each stimulus consisted of a loudspeaker which was placed directly behind a face stimulus (which could be "illuminated" as shown here in a central position in the foreground). For the audio-visual congruent trials, the target was presented from the same location as the crossmodal distractor stimulus. For the audio-visual incongruent trials, the crossmodal distractor stimulus was presented either directly behind (eg central, background position as shown in this example) or in front of the target stimulus. The foreground array was positioned either within peripersonal space in experiments 1 and 2 (ie 60 cm/120 cm for the foreground and background positions, respectively) or in extrapersonal space (80 cm/140 cm) in experiment 3.

The visual stimuli consisted of eight identical copies of an unfamiliar male face (6 cm high × 4 cm wide) printed on acoustically semi-transparent cloth (see McAnally and Martin 2008; Perrott et al 1990 for a similar procedure). Each face stimulus was placed directly in front of a loudspeaker at a distance of approximately 2 cm to ensure the auditory and visual stimuli were spatially congruent. Two 5 volt LEDs were placed behind each face to illuminate it evenly and independently. The visual angles for the face stimuli in the foreground (60 cm) were approximately 6 deg × 4 deg (height × width) and for the faces in the background (120 cm), the visual angles were approximately 3 deg × 2 deg. The entire apparatus was placed in a dark, windowless room.

The stimuli in the background were elevated by 6 cm to ensure that they were not obscured by the stimuli in the foreground. A red fixation light was positioned at a similar height as the audio-visual stimuli and placed on a pole which was positioned directly in front of the participant at a distance of 150 cm. This fixation light remained illuminated throughout the experiment.

Each stimulus was presented for a duration of 200 ms. The same face and voice stimuli were presented from all spatial locations to remove any contextual cues that could be associated with a particular spatial location. The presentation of the stimuli and recording of the timing and accuracy data were controlled by a custom-designed computer programme coded in Visual Basic. Participants responded by pressing either the 'm' or 'z' keys on a computer keyboard to indicate that the target was found in the 'foreground' or 'background' of the array. These response keys were counterbalanced across participants. A target was present in all trials.

2.1.3 *Design*. The experiment was based on a repeated measures design with modality (vision only, auditory only, audio-visual congruent, and audio-visual incongruent) and target angular position (central or peripheral) as the main factors. The experiment was blocked by modality such that there were three blocks for each of the visual-only, auditory-only, and audio-visual (congruent and incongruent) conditions. Block order was counterbalanced across participants. The target was located in either the foreground or background with equal probability. Within each modality block, trials were randomly presented across participants. The dependent variables were response times and accuracy rates.

2.1.4 *Procedure.* There were 80 trials in each unimodal condition and 160 trials in the audio-visual condition. In the audio-visual block, half of the trials included spatially congruent audio-visual stimuli and half included spatially incongruent audio-visual stimuli. In the congruent condition, the auditory and visual stimuli were presented from the same spatial location. In the spatially incongruent trials, the visual and auditory stimuli were presented from different locations in depth but along the same angle (eg the visual stimulus was presented at the left peripheral location in the foreground array, while the auditory stimulus was presented at the left peripheral location in the background array).

In the visual-only and audio-visual conditions, a target face was illuminated, and the participants' task was to determine whether the visual target (ie an illuminated face stimulus) was positioned in the foreground or background of the array by pressing the 'm' or 'z' keys. In the auditory-only condition, the task was to determine whether the sound (voice) emanated from a loudspeaker positioned either in the foreground or background array. None of the faces was illuminated in the auditory-only condition. In the audio-visual condition, participants were told to ignore the auditory stimulus and determine whether the visual target was presented in the foreground or background. In all conditions, participants were told to locate the target as quickly and as accurately as possible, whilst remaining fixated on the central fixation point. Participants were given ten practice trials before each block to ensure they understood the task and the general location of the arrays.

2.2 *Results*

Data from two of the participants were excluded, as accuracy for both participants was less than chance performance. Subsequent analyses were based on data sets from twenty-seven of the twenty-nine participants.

2.2.1 *Accuracy.* In order to compare performance across the spatially congruent and incongruent conditions, performance was averaged across locations in depth (ie across target locations appearing in the foreground or background). Accuracy performance for each of the

modality conditions was: vision-only, 90.98%; auditory-only, 77.15%; audio-visual congruent, 90.47%; and audio-visual incongruent, 86.91%. To determine how congruency affected performance compared to the unimodal conditions, we conducted a 4 × 2 repeated-measures ANOVA with modality (vision-only, audition-only, audio-visual congruent, and audio-visual incongruent) and target angle (central versus peripheral) as factors. There was a main effect of modality ($F_{3,78} = 15.75$, $p < 0.0001$): a posteriori Tukey test on the main effect revealed that performance to the auditory-only condition was significantly worse than performance to the visual-only ($p < 0.002$), audio-visual congruent ($p < 0.002$) and audio-visual incongruent ($p < 0.01$) conditions. There was no effect of target angle ($F_{1,26} < 1$). There was a significant interaction between modality and target angle ($F_{3,78} = 16.52$, $p < 0.0001$), which is depicted in figure 2. A Tukey posteriori analysis of the interaction between modality and target angle revealed the following differences: in the auditory-only condition, performance was worse to the centrally (71.28%) than to the peripherally located target (83.07%; $p = 0.001$) and in the audio-visual incongruent condition, performance was worse to the peripheral location (82.93%) than to the central location (90.895%; $p < 0.006$), but this difference in target angle did not reach significance within any of the other conditions. At the central positions only, performance to the audio-visual congruent trials differed only from that in the auditory-only conditions ($p < 0.001$) and was the same as performance in the vision-only and audio-visual incongruent trials (all ps > 0.8). For the peripheral positions, however, a different pattern emerged. Here, performance to the audio-visual congruent condition was the same as that to the vision-only condition ($p = 0.9$) but was better than to either the auditory-only ($p < 0.02$) or the audio-visual incongruent ($p < 0.02$) conditions.[(1)]

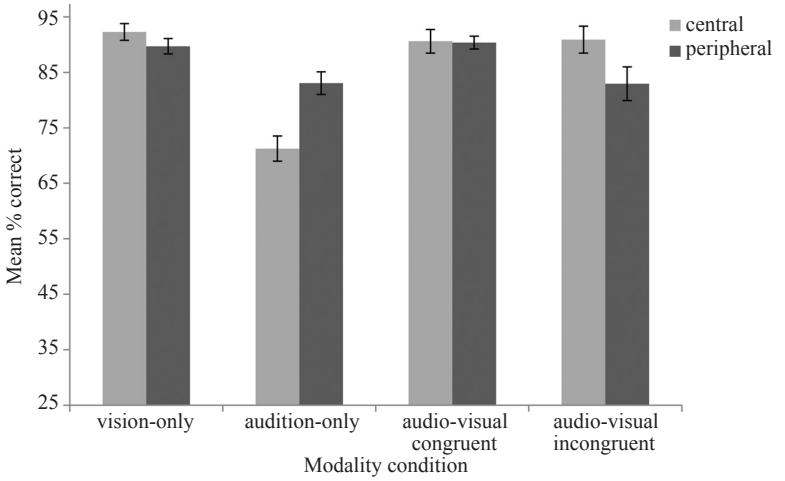

Figure 2. Mean accuracy performance to locating a (visual) target in experiment 1 across each of the modality conditions and target positions. Error bars represent the SEM.

2.2.2 *Reaction time.* A second 4 × 2 repeated-measures ANOVA was performed on the response time data to the unimodal and audio-visual conditions, with modality (vision-only, audition-only, audio-visual congruent, and audio-visual incongruent) and target angle (central versus peripheral) as factors. The difference in response times across the modality conditions failed to reach significance ($F_{3,78} = 1.09$, $p = 0.359$). The mean response times

[(1)] Note that, when target 'Hemifield' (left or right) was also considered as a factor in analysis, we found no evidence for a main effect ($F_{1,26} < 1$); nor did Hemifield affect performance in any of the other conditions in experiment 1. Furthermore, we failed to find evidence of a main effect of Hemifield or interactions with other factors in the following experiments. For clarity, we decided not to include this factor in the main analysis.

to targets in each of the modality conditions were: vision-only, 360 ms; auditory-only, 375 ms; audio-visual congruent, 321 ms; and audio-visual incongruent, 360 ms. The main effect of target angle ($F_{1,26} < 1$) also failed to reach significance, and there was no evidence for an interaction between the factors ($F_{3,78} < 1$). Further analyses (based on the mean response time/mean proportion correct for each participant) on the modality factor confirmed there was no evidence for a speed–accuracy trade-off in the participants' performance.

2.3 *Discussion*

Our results suggest that participants were more accurate at localising a visual target in a depth array compared to an auditory target. However, we found that locating a visual target was less efficient (ie reduced accuracy) when an auditory stimulus was presented at a different location (in front of or behind the visual target in the array) relative to a spatially congruent location, particularly at the more peripheral positions of the array where this difference reached statistical significance. The results of the current study also suggest that spatially congruent auditory stimuli facilitate the localisation of the visual targets at peripheral locations relative to auditory-only targets and to spatially incongruent audio-visual stimuli. However, visual spatial perception did not benefit from congruent multisensory stimulation, since accuracy in the audio-visual congruent condition was the same as in the visual-only condition, which supports the idea of visual dominance in spatial tasks. However, despite this sensory dominance, the results suggest that audition can nevertheless influence the perceived location of the visual target: auditory information presented in a different depth plane to the visual target distracted participants in locating the visual target, particularly when the target appeared in a peripheral location relative to the congruent conditions. In other words, accuracy was reduced when a visual target was presented in the periphery with a spatially incongruent auditory distractor. Therefore, despite the relatively small spatial difference between the visual and auditory stimuli (ie 60 cm), participants were significantly more accurate when audio-visual stimuli were presented in the same location in depth compared to different depths.

Relative to accuracy performance, response times appeared to be unaffected by the modality or by the location of the stimuli. We expected that the presence of an auditory distractor (whether spatially congruent or otherwise) might speed up the orienting of attention towards the general location of the visual target, particularly in the more peripheral positions. However, we found no evidence for this benefit on response times, but we suggest that this may be due to the time involved in subsequently determining the precise location of the visual target (background or foreground) following orienting towards the general location. Further research is required, however, to determine how these processes (orienting towards and localisation of the target) differentially affect response times.

3 Experiment 2

In the previous experiment, spatially incongruent auditory stimuli affected visual target localisation more in the peripheral locations than the central locations. This effect is likely due to auditory spatial information being more reliable in the periphery than visual spatial information (see Bavelier et al 2006 for a review), thus having a greater influence when incongruent with the location of the visual target. The question then arises whether the localisation of peripheral auditory targets is affected by the presence of spatially congruent or incongruent visual information. We would hypothesise that the visual distractors would have little effect on auditory target localisation in the periphery but would have an effect on the central locations (ie the opposite of the results of experiment 1). However, it is also possible that the same results as experiment 1 are attained due to the Colavita

effect (Colavita 1974, 1982; Colavita and Weisberg 1979), whereby visual information dominates the other sensory modality during bimodal presentations. The Colavita effect has been shown to be present even in tasks where visual information is not the most reliable source of information, compared to another modality (eg auditory information in a temporal task—Ngo et al 2010). In the following experiment, we replicated experiment 1, but here the participants' task was to indicate the location of an auditory target.

3.1 *Method*

3.1.1 *Participants.* Thirteen participants (six female) between the ages of 21 and 32 years (mean age = 25.8 years) took part in this experiment. Participants reported having either normal or corrected to normal vision, and all reported normal hearing. None of the participants took part in the previous experiment.

3.1.2 *Design, apparatus, and procedure.* This experiment was identical to experiment 1 with the exception that participants were asked to identify the location of an auditory target as being in the foreground or background of the array (and ignore the visual stimuli) in the audio-visual conditions.

3.2 *Results*

3.2.1 *Accuracy.* A 4 × 2 repeated-measures ANOVA was performed on the accuracy data with modality (vision-only, auditory-only, audio-visual congruent, audio-visual incongruent) and target angle (central versus peripheral) as factors. The mean percentage correct for each of the modalities was visual-only, 91.6%; auditory only, 70.8%; audio-visual congruent, 75.6%; and audio-visual incongruent, 48.7%. There was a main effect of modality ($F_{3,36} = 22.03$, $p < 0.0001$). A posteriori Tukey test revealed that performance to the vision-only condition was significantly better than to all other conditions (all ps < 0.05). Moreover, performance in the audio-visual incongruent condition was worse than that in either the auditory-only or audio-visual congruent (all ps < 0.002) conditions. There was no effect of target angle ($F_{1,12} < 1$, ns). There was a significant interaction between these factors ($F_{3,36} = 2.82$, $p < 0.05$; see figure 3a). A Tukey posteriori did not reveal any significant differences between the central and peripheral locations within each of the modality conditions. However, at the central positions only performance to the audio-visual congruent trials was worse than to the vision-only condition ($p < 0.01$), the same as to the auditory-only condition ($p = 0.94$), and better than to the audio-visual incongruent condition ($p < 0.001$). At the peripheral positions only the same pattern emerged in that performance to the audio-visual congruent trials was worse than the vision-only ($p < 0.02$), the same as auditory-only ($p = 0.93$), and better than the audio-visual incongruent ($p < 0.001$). Importantly, for the audio-visual incongruent condition, performance in locating the auditory target when presented in either the central or peripheral locations was worse than performance to either of these positions in the auditory-only condition.

3.2.2 *Reaction time.* The mean response times to each of the conditions was as follows: vision-only, 436 ms; auditory-only, 642 ms; audio-visual congruent, 682 ms; and audio-visual incongruent, 753 ms (see figure 3b). Once again, we ran a 4 × 2 repeated-measures analysis on the reaction time data between the unimodal and audio-visual conditions. There was a main effect of modality ($F_{3,36} = 7.77$, $p < 0.001$; see figure 3b). A posteriori Tukey test revealed that response times were significantly faster to the vision-only targets than targets presented in either the auditory only ($p < 0.02$), audio-visual congruent ($p < 0.006$), or audio-visual incongruent ($p < 0.004$) conditions. There was no main effect of target angle ($F_{1,12} = 1.83$, $p = 0.20$) and no interaction between these factors ($F_{3,36} = 1.081$, $p = 0.3$). Inverse efficiency analysis (based on the mean response time/mean proportion correct) revealed a main effect of

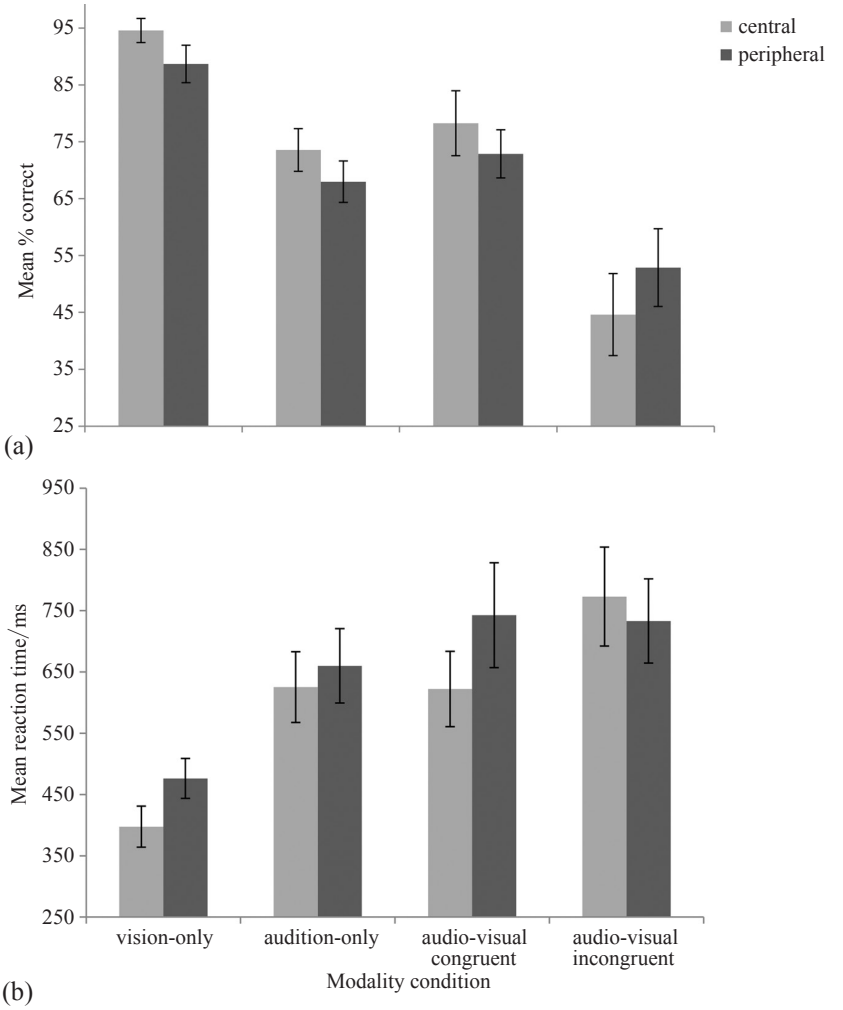

Figure 3. (a) Mean accuracy to locating an (auditory) target in experiment 2 across each of the modality conditions and target positions. (b) Mean reaction times across each of the modality conditions and target positions in experiment 2. Error bars represent the SEM.

modality ($F_{3,36} = 7.06$, $p < 0.001$), which was due to less efficient performance in the audio-visual incongruent condition relative to other conditions. However, on the basis of further analysis, we found no evidence for a speed–accuracy trade-off in participants' performance.

3.3 *Discussion*

We first found that locating a visual-only target was more accurate than locating either an auditory-only target or an auditory target presented together with a visual stimulus. Moreover, we found that participants were more accurate at locating an auditory target when it was presented with a visual distractor from the same spatial location compared to when both were presented from different locations separated in depth. These findings are consistent with the findings on accuracy performance from experiment 1. Again, similar to the findings of experiment 1, the advantage for spatially congruent information was limited to accuracy performance, as participants' reaction times were unaffected by crossmodal spatial congruency (although locating a visual-only target was the fastest).

In contrast to the results from experiment 1, accuracy for locating auditory-only targets in peripheral locations was the same as that to central locations. However, when the auditory target was presented with a visual distractor, performance was better when the visual distractor

was co-located with the auditory target than when it was incongruently located, irrespective of whether these stimuli were presented in central or peripheral positions. In experiment 1, when the target was visual, we observed a benefit for co-located versus incongruently located auditory distractors in the peripheral positions only. In the current experiment, it appears that localising an auditory target in the presence of an incongruent visual distractor was a difficult task. Indeed, performance in this condition was at chance level. One could argue that the difference in the effect of audio-visual incongruent conditions across the central locations in experiments 1 and 2 may be due to the ventriloquist effect (Bertelson et al 2000; Howard and Templeton 1966). In the present experiment the incongruent visual distractor may have captured the perceived location of the auditory target, thus reducing accuracy. Moreover, accuracy in the audio-visual incongruent condition at both central and peripheral locations was around chance level, which was lower than that found in experiment 1, suggesting that the visual stimulus was a more salient distractor than an auditory stimulus, again consistent with the ventriloquist effect. On the other hand, if the ventriloquist effect influenced these results, then we would also expect performance in the condition where the visual distractor was spatially congruent with the auditory target to be more consistent with performance in the vision-only condition, in that the presence of a co-located visual stimulus should enhance the localisation of the auditory target. Although performance in this audio-visual congruent condition was slightly enhanced relative to performance in the auditory-only condition (5% improvement), this difference was not significant. Moreover, performance at locating the visual-only targets was better and faster than in any of the other conditions, including the audio-visual congruent condition. It is possible that because the audio-visual congruent trials were intermixed with the audio-visual incongruent trials, this led to greater uncertainty when localising the auditory target in the incongruent condition. However, a similar, general cost on performance in the incongruent condition in experiment 1 was not found. Our results likely suggest important interactions between vision and audition at all locations of the target in the array, dependent on the modality of the target and the crossmodal distractor stimuli.

4 Experiment 3

In experiments 1 and 2, the foreground array was located within the participant's reach, ie in peripersonal space, while the background array was located in extrapersonal space. Previous research has suggested that there are qualitative differences between the perception of objects positioned within peripersonal space and those positioned in extrapersonal space (Farnè and Làdavas 2002). For example, Farnè and Làdavas found that a distance of 70 cm (and beyond) between the participant and an object was sufficient to produce audio-tactile extinction effects associated with extrapersonal space (Farnè and Làdavas 2002). They argued that objects located in peripersonal space are more likely to draw the attention of the observer than are objects in extrapersonal space, since objects that are close are most likely to cause bodily harm and thus require relatively greater allocation of attention. Although their study focused on the audio-haptic interactions in which the participant was required to perform an action on an object, such as grasping the object (Cant and Goodale 2005, 2006), little is known about whether these distinctions between peripersonal and extrapersonal space also apply to audio-visual interactions. We might, however, expect that the perception of a person's face or verbal greeting may be affected by the relative distance between the other person and oneself (see Burgoon 1978).

The following experiment was designed, therefore, to investigate the role of stimulus distance from the observer on crossmodal interactions in a target localisation task. Here, we replicated experiment 1; however, the entire apparatus was moved a further 20 cm away from the participant into extrapersonal space (ie beyond arm's reach). With this manipulation,

we investigated whether the relative benefit on target localisation performance from a spatially congruent crossmodal distractor was affected by the distance of the targets from the observer.

4.1 *Method*

4.1.1 *Participants.* Sixteen participants (nine female) between the ages of 19 and 27 years (mean age was 22.06 years) took part in this experiment. All reported to have normal or corrected-to-normal vision and did not report any hearing impairments. All were naive to the purposes of the experiment and none took part in experiments 1 and 2.

4.1.2 *Design, apparatus and procedures.* The apparatus was identical to experiment 1, with the exception that the locations of the stimulus array were now placed at distances of 80 cm (foreground array) and 140 cm (background array) from the participant. Thus, both the foreground and background stimulus arcs were positioned outside the participant's reachable (ie peripersonal) space, but the 60 cm separation between the two arcs was maintained as in the previous experiment. The visual angles for the face stimuli in the foreground were 4.3 deg × 2.9 deg (height × width), and for the faces in the background the visual angles were approximately 2.5 deg × 1.6 deg.

The design and procedures were the same as in experiment 1: the participants' tasks were to determine the location of the relevant target in the unimodal conditions and the visual target in the crossmodal conditions (whilst ignoring the auditory stimulus).

4.2 *Results*

4.2.1 *Accuracy.* The mean accuracy performance across each condition was as follows: vision-only, 90.65%; auditory-only, 63.16%; audio-visual congruent, 90.39%; and audio-visual incongruent, 87.03%. A 4 × 2 repeated-measures analysis (as in experiments 1 and 2) was performed on the accuracy data. There was a main effect of modality ($F_{3,45} = 121.29$, $p < 0.0001$). A posteriori Tukey test revealed that accuracy to the auditory-only condition was lower than accuracy in either the visual-only ($p < 0.001$), audio-visual congruent ($p < 0.001$), or audio-visual incongruent ($p < 0.001$) conditions. None of the other differences across the modality conditions reached significance. The effect of target angle failed to reach significance ($F_{1,15} = 3.06$, $p = 0.10$).

There was a significant interaction between these factors ($F_{3,45} = 21.09$, $p < 0.0001$) as shown in figure 4a. A posteriori Tukey analysis revealed significantly better accuracy in locating a centrally presented target (93.72%) than a peripherally presented target (87.58%) in the vision-only condition ($p = 0.02$) and in the audio-visual incongruent condition (central = 91.09%; peripheral = 82.97%; $p = 0.001$). Performance across positions in the audio-visual congruent condition (central = 93.13%; peripheral = 87.66%) was equivalent ($p = 0.06$). Conversely, auditory-only peripheral targets (67.91%) were more accurately located than centrally (58.42%) presented targets ($p < 0.0001$).

At central locations only, performance to the audio-visual congruent condition was the same as that to the visual-only and audio-visual incongruent conditions. Performance to the audio-visual congruent condition was better than to the auditory-only condition at the central locations ($p < 0.001$). At peripheral locations only, performance to the audio-visual congruent condition was the same as that to the vision-only condition, better than that to the auditory-only condition ($p < 0.001$) but failed to reach statistical difference to the audio-visual incongruent condition ($p = 0.164$).

4.2.2 *Reaction time.* The mean response times to each of the conditions were as follows: vision-only, 366 ms; auditory-only, 506 ms; audio-visual congruent, 339 ms; audio-visual incongruent, 354 ms. As before, a 4 × 2 ANOVA was performed on reaction time performance.

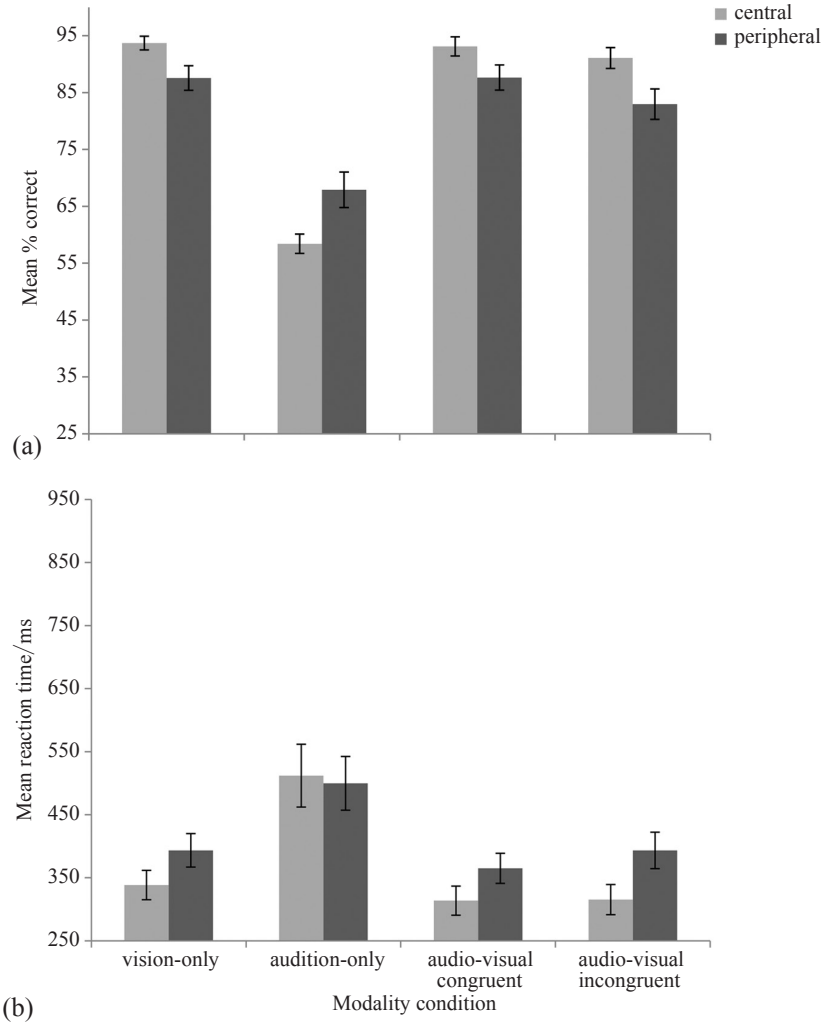

Figure 4. (a) Mean accuracy and (b) mean response times to locating a (visual) target in experiment 3 across each of the modality conditions and target positions. The arrays were positioned at 80 cm and 140 cm, relative to the participant. Error bars represent the SEM.

There was a main effect of modality ($F_{3,45} = 17.80$, $p < 0.0001$): a posteriori Tukey test revealed slower response times to the auditory-only targets than to targets in either the visual-only ($p < 0.001$), audio-visual congruent ($p < 0.001$), or audio-visual incongruent ($p < 0.001$) conditions. None of the other pairwise comparisons reached significance. There was also a main effect of target angle ($F_{1,15} = 41.02$, $p < 0.001$), with faster responses to centrally (370 ms) than to peripherally (412 ms) located targets.

There was a significant interaction between these factors ($F_{2,45} = 10.49$, $p < 0.001$), as shown in figure 4b. A posteriori Tukey analysis revealed that response times were significantly slower (all ps < 0.01) for peripheral targets compared to central targets in the vision-only (central, 338 ms; peripheral, 393 ms), audio-visual congruent (central, 313 ms; peripheral, 365 ms), and audio-visual incongruent (central, 315 ms; peripheral, 393 ms) conditions, but there was no difference found between target locations in the auditory-only condition (511 ms and 499 ms respectively).

Response times to the centrally located targets in the audio-visual congruent condition were the same as those to the centrally located targets in other conditions (except for in the auditory-only condition). Response times to the peripherally located targets in the

audio-visual congruent condition were the same as those to targets in either the vision-only or audio-visual incongruent only (although they were faster than those to auditory-only targets). An inverse efficiency analysis (based on the mean response time/mean proportion correct for each participant) revealed a main effect of modality ($F_{3,45} = 49.8$, $p < 0.001$), which was due to less efficient performance in the auditory-only condition relative to other conditions. In general, on the basis of further analysis, we found no evidence for a speed–accuracy trade-off in participants' performance.

4.2.3 *Comparison across distances.* To determine whether performance was affected by the distance of the stimulus array from the participant (ie peripersonal or extrapersonal space), we conducted a 3-way mixed-design ANOVA with array distance (experiment 1 versus experiment 3) as the between-subjects factor and congruency (audio-visual congruent versus audio-visual incongruent) and target angle (central versus peripheral) as the within-subject factors on both the accuracy and response time data. We found no effect of distance on either the accuracy data ($F_{1,37} < 1$, ns) or the response times ($F_{1,37} < 1$, ns). Moreover, distance did not interact with any of the other factors for either the accuracy or response time data. In other words, performance in localising audio-visual stimuli was not qualitatively affected by the distances between the spatial array and the participant. The effect of congruency was significant for accuracy performance only ($F_{1,37} = 7.65$, $p = 0.009$) but not for response times ($F_{1,37} = 1.83$, $p = 0.18$). A posteriori test revealed more accurate performance in locating visual targets when presented with a congruently located (90.54%) than an incongruently located (85.61%) auditory stimulus in depth. There was also a main effect of target angle with better and faster performance to centrally located (90.62%; 307 ms) than peripherally located (85.52%; 369 ms) targets ($F_{1,37} = 9.58$, $p = 0.004$ and $F_{1,37} = 45.52$, $p < 0.0001$ for accuracy and response times, respectively). Finally, for the accuracy data only, we found an interaction between audio-visual congruency and target angle ($F_{1,37} = 7.96$, $p = 0.008$). Performance was worse when a visual target was presented in the periphery with a spatial incongruent auditory stimulus (81.60%) compared to when the visual target was presented in the periphery with a congruent auditory stimulus (89.45%; $p < 0.001$). No other comparisons reached significance.

4.3 *Discussion*

Participants were significantly more accurate and faster at locating the target in the vision-only condition, and in both audio-visual conditions compared to the auditory-only condition. As in experiment 1, localising a visual target along a depth plane was marginally more efficient when the auditory stimulus was co-located than when it was spatially incongruent, although this was not statistically different. However, similar to experiment 1, these differences were observed in the peripheral locations only, with a 5% drop in accuracy performance and 30 ms slowing of response times when the location of the sound was incongruent to the peripheral location of the visual target relative to when the sound was congruently located (although these differences failed to reach statistical significance). Moreover, the location of an incongruent sound had a larger effect on locating a visual target in the periphery than in the centre. On the other hand, when the visual target was centrally located, there was no evidence of a relative benefit on target localisation with a spatially congruent auditory distractor.

It is unclear why we did not get a larger advantage for the audio-visual congruent stimuli over the audio-visual incongruent stimuli, particularly at the peripheral locations, as we found in experiment 1. It may be the case that as the target array is moved further away from the participant the advantage in localising centrally positioned visual targets over peripherally localised targets when the auditory distractor is incongruent becomes less distinct. However, when we compared performance across array distances (ie experiments 1 and 2), no such distinction was found. It is important to note that we did find a main effect of congruency

when comparing experiments 1 and 3. This suggests that the same trend of performance exists for both array distances.

If attention played a role in target detection, we would have expected reaction times for the peripersonal array (experiment 1) to be faster than the responses for targets located in extrapersonal space. However, the results of experiment 3 suggest that the distance between the loudspeaker/visual stimulus array and the participant did not affect how participants integrated audio-visual information along the depth plane. We add, however, that these findings may be particular to the distances tested here, and that it may be the case that performance would change significantly with greater distance between the participant and the array.

5 General discussion

The importance of spatial congruency across vision and auditory stimuli located in depth when locating a visual target is clearly illustrated by these results. Although spatially congruent audio-visual stimuli did not improve performance over-and-above vision alone, performance in locating a target when visual and auditory stimuli were incongruent was reduced when identifying the location of either a visual (experiment 1) or an auditory (experiment 2) target. Although in experiments 1 and 2 stimuli were presented in an array which extended 60 cm to 120 cm in front of the participant, the effects of the incongruently located crossmodal stimuli were evident regardless of the relative proximity of the stimulus array to the participant (experiment 3).

The effect of incongruently located auditory and visual stimuli is isolated to the more peripheral rather than the central locations, and there was little evidence of crossmodal interference when the target was presented at a central position in the array. The auditory stimuli reduced target localisation performance for visual targets located in the peripheral locations where spatial identification of the visual targets was more difficult (De Valois and De Valois 1988; Rayner and Pollatsek 1992), as seen in the performance to the visual-only conditions. However, the relative lack of an effect when an incongruently presented audio-visual stimulus is presented at a central location suggests that participants could effectively ignore the auditory stimulus when the visual information was relatively reliable. However, for locations where the visual target was more difficult to locate, such as in the periphery, the auditory information had more influence. The results of experiment 2, in which the task was to locate an auditory stimulus, are consistent with this idea: when participants were asked to identify the location of an auditory target, there was no difference in performance to the audio-visual incongruent trials when the auditory target was located either centrally or peripherally. This finding suggests that visual information affected the perceived location of the auditory target in the same fashion across the central and peripheral locations.

In general, our effects were largely limited to accuracy performance, rather than response times. As mentioned earlier, it is possible that the orienting of attention to the location of a target along the azimuth facilitates response times, but then the process of establishing the precise location of the target in depth may be a relatively slower process. Future research is required to establish whether these are independent processes and to tease apart the relative influence of these effects.

Our findings generally extend the previous findings of interactions between vision and audition in spatial perception (Arnott and Goodale 2005; Driver and Spence 1998; Razavi et al 2007) and specifically extend previous findings on aurally aided visual search (Perrott et al 1995, 1996). Such studies typically show better accuracy and faster reaction times when auditory and visual stimuli are presented from the same relative to different locations across the horizontal plane. In contrast to our study, however, Perrott and colleagues reported evidence for a multisensory facilitation such that locating a visual target when an auditory stimulus was spatially congruent resulted in significantly better performance than detecting

the target through vision alone. Our results, on the other hand, suggest that localising a visual target is not facilitated by the presence of a spatially congruent auditory stimulus, relative to vision alone, but that a spatially incongruent auditory stimulus induced a cost in performance, particularly in the periphery.

Although a benefit for congruently located auditory and visual stimuli on performance was not found here, it is important to note that there is a difference between distance and location perception (Abrams and Landgraf 1990). Previous studies on visual distance perception found little improvement on performance by combining audio-visual information (see Loomis et al 1999, 2002) while others found some improvement in auditory distance perception when visual information is provided (Zahorik 2001; Zahorik et al 2005). In the experiments presented here, participants were not explicitly asked to indicate the distance between the stimuli or between the stimuli and themselves. The purpose of this study was to investigate whether spatial congruency across the senses can affect target detection along the depth plane in the same fashion as along the horizontal plane. Therefore, the location-based effects shown here may not occur if participants were given a more explicit distance judgment task (eg indicate the distance between yourself and the target).

Acknowledgments. This research was funded by grant no. 06/IN.1/I96 from Science Foundation Ireland (awarded to FNN).

References

Abrams R A, Landgraf J Z, 1990 "Differential use of distance and location information for spatial localization" *Perception & Psychophysics* **47** 349–359

Alais D, Burr D, 2004 "No direction-specific bimodal facilitation for audiovisual motion detection" *Cognitive Brain Research* **19** 185–194

Alais D, Carlile S, 2005 "Synchronizing to real events: Subjective audiovisual alignment scales with perceived auditory depth and speed of sound" *Proceedings of the National Academy of Sciences of the USA* **102** 2244–2247, doi:10.1073/pnas.0407034102

Arnott S R, Goodale M A, 2005 "Distorting visual space with sound" *Journal of Vision* **5**(8) 172

Bavelier D, Dye M W G, Hauser P C, 2006 "Do deaf individuals see better?" *Trends in Cognitive Sciences* **10** 512–518, doi:10.1016/j.tics.2006.09.006

Bertelson P, Gelder B de, 2003 "The psychology of multimodal perception", in *Crossmodal Space and Crossmodal Attention* Eds C Spence, J Driver (Oxford, UK: Oxford University Press) pp 140–177

Bertelson P, Vroomen J, Gelder B de, Driver J, 2000 "The ventriloquist effect does not depend on the direction of deliberate visual attention" *Perception & Psychophysics* **62** 321–332

Burgoon J K, 1978 "A communication model of personal space violations: Explication and an initial test" *Human Communication Research* **4** 129–142

Cant J S, Goodale M A, 2005 "An fMRI investigation of the perception of form, texture, and colour in human occipito-temporal cortical pathways" *Journal of Vision* **5**(8) 245

Cant J S, Goodale M A, 2006 "Attention to form or surface properties modulates different regions of human occipitotemporal cortex" *Cerebral Cortex* **17** 713–731

Colavita F B, 1974 "Human sensory dominance" *Perception & Psychophysics* **16** 409–412

Colavita F B, 1982 "Visual dominance and attention in space" *Bulletin of the Psychonomic Society* **19** 261–262

Colavita F B, Weisberg D, 1979 "A further investigation of visual dominance" *Perception & Psychophysics* **25** 345–347

Coleman P D, 1963 "An analysis of cues to auditory depth perception in free space" *Psychological Bulletin* **60** 302–315

De Valois R L, De Valois K K, 1988 *Spatial Vision* (New York: Oxford University Press)

Driver J, Spence C, 1998 "Attention and the crossmodal construction of space" *Trends in Cognitive Sciences* **2** 254–262

Farnè A, Làdavas E, 2002 "Auditory peripersonal space in humans" *Journal of Cognitive Neuroscience* **14** 1030–1043, doi:10.1162/089892902320474481

Fukusima S S, Loomis J M, Da Silva J A, 1997 "Visual perception of egocentric distance as assessed by triangulation" *Journal of Experimental Psychology: Human Perception and Performance* **23** 86–100

Gardner M B, 1968 "Proximity image effect in sound localization" *Journal of the Acoustical Society of America* **43** 163

Howard I P, Templeton W B, 1966 *Human Spatial Orientation* (Oxford: John Wiley & Sons)

Jackson C V, 1953 "Visual factors in auditory localization" *Quarterly Journal of Experimental Psychology* **5** 52–65

Kearney G, Gorzel M, Boland F, Rice H, 2010 "Depth perception in interactive virtual acoustic environments using higher order ambisonic sounds", paper presented at the 2nd International Symposium on Ambisonics and Spherical Acoustics, Paris, France

Loomis J M, Da Silva J A, Fujita N, Fukusima S S, 1992 "Visual space perception and visually directed action" *Journal of Experimental Psychology: Human Perception and Performance* **18** 906–921

Loomis J M, Klatzky R J, Golledge R G, 1999 "Auditory distance perception in real, virtual, and mixed environments", in *Mixed Reality: Merging Real and Virtual Worlds* Eds Y Ohta, H Tamura (Tokyo: Ohmsha) pp 201–214

Loomis J M, Klatzky R L, Philbeck J W, Golledge R G, 1998 "Assessing auditory distance perception using perceptually directed action" *Perception & Psychophysics* **60** 966–980

Loomis J M, Lippa Y, Klatzky R J, Golledge R G, 2002 "Spatial updating of locations specified by 3-D sound and spatial language" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **28** 335–345

McAnally K, Martin R, 2008 "Sound localisation during illusory self-rotation" *Experimental Brain Research* **185** 337–340

McDonald J J, Teder-Sälejärvi W A, Hillyard S A, 2000 "Involuntary orienting to sound improves visual perception" *Nature* **407** 906–908

Ngo M K, Sinnett S, Soto-Faraco S, Spence C, 2010 "Repetition blindness and the Colavita effect" *Neuroscience Letters* **480** 186–190, doi:10.1016/j.neulet.2010.06.028

Perrott D R, 1984 "Discrimination of the spatial distribution of concurrently active sound sources: Some experiments with stereophonic arrays" *Journal of the Acoustical Society of America* **76** 1704–1712

Perrott D R, Cisneros J, McKinley R L, D'Angelo W R, 1995 "Aurally aided detection and identification of visual targets" *Human Factors and Ergonomics Society Annual Meeting Proceedings* **39** 104–108

Perrott D R, Cisneros J, McKinley R L, D'Angelo W R, 1996 "Aurally aided visual search under virtual and free-field listening conditions" *Human Factors* **38** 702–715

Perrott D R, Saberi K, Brown K, Strybel T Z, 1990 "Auditory psychomotor coordination and visual search performance" *Perception & Psychophysics* **48** 214–226

Philbeck J W, Loomis J M, 1997 "Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions" *Journal of Experimental Psychology: Human Perception and Performance* **23** 72–85

Plumert J M, Kearney J K, Cremer J F, 2004 "Distance perception in real and virtual environments", paper presented at the Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization, Los Angeles, CA

Rayner K, Pollatsek A, 1992 "Eye movements and scene perception" *Canadian Journal of Psychology* **46** 342–376

Razavi B, O'Neill W E, Paige G D, 2007 "Auditory spatial perception dynamically realigns with changing eye position" *Journal of Neuroscience* **27** 10249–10258

Sekuler R, Sekuler A B, Lau R, 1997 "Sound alters visual motion perception" *Nature* **385** 308

Shams L, Kamitani Y, Shimojo S, 2000 "What you see is what you hear" *Nature* **408** 788

Shelton B R, Searle C L, 1980 "The influence of vision on the absolute identification of sound-source position" *Attention, Perception, & Psychophysics* **28** 589–596, doi: 10.3758/bf03198830

Spence C, 2007 "Audiovisual multisensory integration" *Acoustical Science and Technology* **28** 61–70

Spence C, Driver J, 2000 "Attracting attention to the illusory location of a sound: reflexive crossmodal orienting and ventriloquism" *NeuroReport* **11** 2057–2061

Sugita Y, Suzuki Y, 2003 "Audiovisual perception: Implicit estimation of sound-arrival time" *Nature* **421** 911

Van der Burg E, Olivers C N L, Bronkhorst AW, Theeuwes J, 2008 "Pip and pop: nonspatial auditory signals improve spatial visual search" *Journal of Experimental Psychology: Human Perception and Performance* **34** 1053–1065

Warren D H, 1970 "Intermodality interactions in spatial localization" *Cognitive Psychology* **1** 114–133, doi:10.1016/0010-0285(70)90008-3

Welch R B, DuttonHurt L D, Warren D H, 1986 "Contributions of audition and vision to temporal rate perception" *Perception & Psychophysics* **39** 294–300

Welch R B, Warren D H, 1986 "Intersensory interactions", in *Handbook of Perception and Human Performance* Eds K R Boff, L Kaufman, J P Thomas (Chichester, Sussex: Wiley-Interscience) pp 25.21–25.36

Zahorik P, 1998 *Experiments in Auditory Distance Perception* PhD thesis, University of Wisconsin, Madison

Zahorik P, 2001 "Estimating sound source distance with and without vision" *Optometry and Vision Science* **78** 270–275

Zahorik P, 2002 "Auditory display of sound source distance", paper presented at the International Conference on Auditory Display, Kyoto, Japan

Zahorik P, Brungart D S, Bronkhorst A W, 2005 "Auditory distance perception in humans: a summary of past and present research" *Acta Acustica* **91** 409–420

Zampini M, Shore D I, Spence C, 2003 "Audiovisual temporal order judgments" *Experimental Brain Research* **152** 198–210

Ziemer C J, Plumert J M, Cremer J F, Kearney J K, 2009 "Estimating distance in real and virtual environments: Does order make a difference?" *Attention, Perception, & Psychophysics* **71** 1095–1106, doi:10.3758/app.71.5.1096