

# Spatially uninformative sounds increase sensitivity for visual motion change

Sabine M. Staufenbiel · Rob H. J. van der Lubbe · Durk Talsma

Received: 29 April 2011 / Accepted: 8 July 2011 / Published online: 30 July 2011  
© The Author(s) 2011. This article is published with open access at Springerlink.com

**Abstract** It has recently been shown that spatially uninformative sounds can cause a visual stimulus to pop out from an array of similar distractor stimuli when that sound is presented in temporal proximity to a feature change in the visual stimulus. Until now, this effect has predominantly been demonstrated by using stationary stimuli. Here, we extended these results by showing that auditory stimuli can also improve the sensitivity of visual motion change detection. To accomplish this, we presented moving visual stimuli (small dots) on a computer screen. At a random moment during a trial, one of these stimuli could abruptly move in an orthogonal direction. Participants' task was to indicate whether such an abrupt motion change occurred or not by making a corresponding button press. If a sound (a short 1,000 Hz tone pip) co-occurred with the abrupt motion change, participants were able to detect this motion change more frequently than when the sound was not present. Using measures derived from signal detection theory, we were able to demonstrate that the effect on accuracy was due to increased sensitivity rather than to changes in response bias.

**Keywords** Multisensory integration · Motion · Attentional capture · Accuracy · Visual search

## Introduction

In everyday life, we receive and process a host of sounds, images, and other sensations. Each of these sensations alone often provides insufficient and ambiguous information about the environment we live in. Combined with each other, however, they may lead to a coherent interpretation of this environment. Investigations into the processes involved in the integration of information across the senses have recently led to many new insights into the way our perceptual systems work.

For instance, it has been suggested that visual and auditory stimuli when presented in spatial alignment within a short time window are perceived as more salient and less ambiguous than when they are presented in isolation (Stein et al. 1996; Calvert et al. 2000). Single cell recordings in animal superior colliculus neurons have shown that these multisensory integration processes are related to enhanced neural firing patterns (Stein et al. 2005). These enhanced firing patterns are consistent with a host of behavioral findings showing that detection thresholds for multisensory stimuli are lower than those for unisensory stimuli (Stein et al. 1996; Frassinetti et al. 2002; Noesselt et al. 2008).

Vroomen and De Gelder (2000) have reported further evidence for the notion that visual stimuli are more easily detectable when accompanied by a sound. These authors used rapidly and randomly changing patterns of four dots. Occasionally, the four dots were presented in a pre-defined configuration (i.e., outlining a diamond shape) and participants had to indicate, by means of a four choice button press, in which corner of the display this particular dot

---

S. M. Staufenbiel · R. H. J. van der Lubbe · D. Talsma  
Department of Cognitive Psychology and Ergonomics,  
University of Twente, Enschede, The Netherlands

R. H. J. van der Lubbe  
Department of Cognitive Psychology, University of Finance  
and Management, Warsaw, Poland

D. Talsma (✉)  
Department of Experimental Psychology, Faculty of Psychology  
and Educational Sciences, Ghent University,  
Henri Dunantlaan 2, 9000 Ghent, Belgium  
e-mail: durktals@gmail.com

configuration occurred. A key finding in this study was that participants were better at detecting the target configurations when a relatively unique sound was presented at about the same moment as the target display. This effect was referred to as the “freezing” phenomenon due to the fact that the sounds appear to freeze the visual configurations for a brief period of time. The freezing phenomenon thus provides support for the notion that sounds can decrease the detection thresholds for visual stimuli, even when the physical locations of the visual and auditory stimuli are not perfectly aligned. Subsequent work (Van der Burg et al. 2008, 2011; Ngo and Spence 2010), using visual search tasks, has shown that a visual target stimulus that was not very salient by itself could become instantly noticeable when it was accompanied by a short tone. This result, labeled the “Pip and Pop” effect, suggests that multisensory stimuli are indeed able to capture attention and therefore that multisensory integration processes themselves operate pre-attentively.

Although many recent studies have addressed the question how multisensory integration processes interact with attention (Talsma et al. 2009; Soto-Faraco et al. 2004; Soto-Faraco and Alsius 2007; Navarra et al. 2010; Van der Burg et al. 2009; Van Ee et al. 2009; Alais et al. 2010; Bertelson et al. 2000; Vroomen et al. 2001a, b), it still remains unclear to what degree the influence of multisensory integration is limited to specific forms of attentional feature selection, such as the selection of color and orientation, or whether this influence can be extended to other forms of attentional selection. The goal of the present paper is to address this question.

Girelli and Luck (1997), using a visual search task, demonstrated that target stimuli that were either defined by motion, color, or orientation, activate a common attentional mechanism. More specifically, using a visual search task, in which a singleton stimulus could be present that was defined by color, shape, or motion, they showed that an event-related potentials (ERP) component known as the N2pc was similar regardless of which stimulus dimension defined the singleton. The authors noted that this finding is consistent with the known anatomy and physiology of the visual system. This finding supports the general view that the properties of the individual objects are analyzed by a common set of structures within the ventral pathway, independent of the specific features being processed (Goodale and Milner 1992).

Burr and Alais (2006) indicated that the superior colliculus has strong reciprocal links with the middle temporal (MT) cortical area (Standage and Benevento 1983), which is an area specialized for the processing of visual movement and whose activity is strongly correlated with visual motion perception (Britten et al. 1992). Burr and Alais further summarized several studies, showing that MT

projects directly to the ventral intra-parietal area. In the latter area, bimodal cells combine the motion-related input from area MT with input from auditory areas (Bremmer et al. 2001; Graziano 2001), again suggesting that sound may affect motion perception.

Taking into account these close ties between audiovisual- and motion-related processing mechanisms, we hypothesized that the findings of Van der Burg et al. (2008) could be extended to moving stimuli. It should be noted that current evidence with regard to an auditory influence on moving stimuli is somewhat mixed, however. Using speech stimuli, Alsius and Soto-Faraco (2011) did not observe that auditory stimuli influenced the detection of lip movements. Likewise, Fujisaki et al. (2006) used rotating visual stimuli and an amplitude-modulated continuous auditory signal. In their task, participants were required to detect which visual stimulus rotated at the same frequency as the modulation frequency of the tone. Although participants were able to detect this, they failed to do so in the highly automated fashion reported by Van der Burg et al.

Several recent studies show that sound can influence visual motion processing (Calabro et al. 2011; Roseboom et al. 2011; Arrighi et al. 2009). For instance, Arrighi et al. (2009) reported that the detection of a specific form of biological motion, namely tap dancing can be facilitated using meaningful auditory cues (the tapping rhythm of the dancer’s movement). Calabro et al. (2011) reported that spatially co-localized directionally congruent auditory stimuli can enhance the detection of visual motion processing. In summary, the current literature shows that under certain circumstances auditory stimuli can enhance the detection of visual motion stimuli, but it remains the question how, and under what circumstances these interactions take place. Importantly, no study to date has shown that spatially uninformative sounds can improve the detection of motion-related changes in a way comparable to the effect described by Van der Burg et al.

The goal of the current study was to address the question whether such spatially uninformative auditory stimuli can improve the detection of visual motion changes. This was done by employing a detection task in which the participants had to observe an array of dots moving in random directions on a computer display. The participants were instructed to focus on the visual stimuli and to detect a motion direction change of one of the dots (occurring on 50% of all trials). On 50% of the trials in which the motion change occurred, as well as on 50% of the trials on which the motion change did not occur, a sound was presented. If a motion direction change occurred on such a trial, the sound coincided with this direction change.

We hypothesized that participants would be able to detect motion changes among a higher number of distractor items when the sound was present than when the sound was

absent. To investigate this hypothesis (experiment 1a), we used a tracking algorithm that regulated the number of moving visual items in each trial (see also Kim et al. 2010 for a similar algorithm). The number of displayed stimuli increased or decreased depending on the performance of each participant to ensure a stable accuracy of 80% for each participant. Additionally, we expected that participants' sensitivity to direction changes would be higher when a sound was present than when the sound was absent. To test this latter hypothesis (experiment 1b), we presented each participant with a constant number of stimuli and computed each participants' sensitivity to the motion change by computing the sensitivity index ( $d'$ ) and the likelihood ratio ( $\beta$ ) parameters for the sound-present and sound-absent conditions, respectively. To foreshadow our results, participants were indeed better at detecting motion direction changes of visual stimuli when they were accompanied by a sound than when they were presented in silence. The latter effect was expressed in a higher number of visual stimuli that participants could keep track of (experiment 1a) and in a higher  $d'$  value (experiment 1b) when sounds were presented.

## Methods

### Participants

Fourteen participants (13 women, mean age 22.7, ages between 17 and 24), took part in this study. All subjects were students at the University of Twente and received course credits for participation. All participants had normal hearing functions and normal or corrected-to-normal vision. They gave written informed consent and were naïve as to the purpose of the experiment. The study was approved by the ethics committee of the Faculty of Behavioral Sciences at the University of Twente.

### Task and stimuli

Randomly moving white dots were presented on a black background. On 50% of the trials, one of these dots changed direction at an angle of  $90^\circ$  during the trial (direction-change trials), whereas on the other 50% of the trials all dots continued to move in their original direction (continuous-motion trials). Additionally, on 50% of the direction-change trials, as well as on 50% of the continuous-motion ones, a short sound was presented (sound-present trials). On the direction-change trials, the sound would appear immediately after (i.e., starting directly after the first frame at which the dot moved in a new direction) the onset of the change in motion path, whereas on the

continuous-motion trials this sound was presented at a comparable point in time without a motion direction change. Notably, the sound provided no information about the location, the presence of a motion direction change, or the direction of the motion change of the dots. On the other 50% of all trials, the sound was absent (sound-absent trials). Participants were instructed to keep their eyes fixed at a red dot located at the center of the screen. At the end of each trial, participants had to respond by making an unsped-up button-press response. They had to press the 'j' key (for "ja", the Dutch equivalent for "yes") of the computer's keyboard if they thought a direction change had occurred and they had to press the 'n' key (for "nee", the Dutch equivalent for "no") if they thought that no direction change occurred.

The above-described task was administered using two different versions. In the first version (experiment 1a), the total number of dots presented on each trial was dynamically updated on the basis of each participant's accuracy using a staircase algorithm. Starting with an initial number of four dots, one dot was to be removed after an error had been made and three dots were to be added after five consecutive correct responses had been given. This procedure was restricted so that a minimum of two dots would always remain present. This procedure ensured that participants would maintain an overall accuracy of about 80%. The number of dots was tracked separately for the sound-present and sound-absent trials. Consequently, the mean number of dots presented in each condition (sound absent or sound present) is the main dependent measure in experiment 1a.

In the second version, (experiment 1b), the basic procedure was the same as that of experiment 1a, except for the fact that the number of visual moving objects was individually adjusted based on participants performance in experiment 1a. More specifically, the average number of objects, averaged across all trial blocks of experiment 1a and rounded off to the nearest integer was used in experiment 1b.

### Procedure

After completion of the informed consent forms, participants received task-specific instructions and then completed four practice blocks of experiment 1a: two with and two without sound stimuli. The order of blocks was distributed randomly. Each block contained twenty trials. Subsequently, participants performed another six blocks of trials for experiment 1a, consisting of three blocks with sound and three blocks without sound. Upon completion of experiment 1a, participants were given a brief pause and then completed experiment 1b (also consisting of six blocks of trials, 3 with sound, and 3 without sound).

## Apparatus and stimuli

The experiment was conducted in a sound attenuated, dimly lit room. Stimulus presentation, timing, and data collection were achieved by using the E-prime 1.1 experimental software package on a standard Pentium® IV class PC. Stimuli were presented on a 17 Inch Philips 107T5 display running at 800 by 600 pixel resolution in 32 bit color and refreshing at a rate of 60 Hz. The viewing distance was approximately 60 cm. Input was given by means of a standard computer keyboard.

Each trial consisted of three phases. First, at least 120 image frames, each lasting 16.7 ms, were presented before the direction change could take place. Secondly, a variable number of frames were presented; on each of these frames, there was a likelihood of 16% that a motion direction change could be programmed. Once the direction change was programmed, this sequence terminated. Thirdly, there were another 120 frames before the trial was finished. To keep the timing characteristics between direction-change and continuous-motion trials identical, a motion direction change was programmed for the continuous-motion trials, according to the above-described procedure, but not executed. In contrast, on the direction-change trials, a motion direction change was both programmed and executed.

Whenever a dot reached the edge of the screen, it bounced back. The spontaneous direction change that participants were required to detect was restricted to take place within a bounding box (defined as 1/8 to 7/8 of the horizontal and vertical screen dimensions). This was done to prevent the target stimulus from being mistaken with the bouncing dot occurring at the edge of the screen. To prevent the stimuli from moving continuously outside of the bounding box, the initial direction of all the dots was limited so that they could not move horizontally or vertically but only diagonally (with a random deviation of  $\pm 15^\circ$  from each diagonal). The direction change of the target stimulus had an angle of  $90^\circ$  with respect to its original trajectory.

Each dot had a radius of three pixels and moved with a velocity of one pixel ( $0.036^\circ$ ) per frame (equaling to  $2.16^\circ/s$ ). The sound used in this experiment was a 15 ms, 1,000 Hz sine wave and included a 5 ms fade-in as well as a 5 ms fade-out to eliminate any audible transients. It was presented via a single loudspeaker directly in front of the participant.

## Data analysis

For experiment 1a, condition-wise mean accuracies and the mean number of on-screen visual objects were calculated for each participant. These data were then analyzed using pair-wise *t* tests. This was done separately for accuracy and mean number of objects. It should be noted that the mean

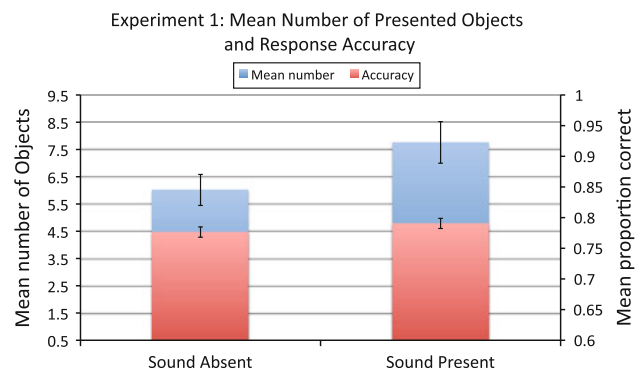
number of on-screen object was the main dependent variable in experiment 1a.

For experiment 1b, the mean response accuracy was calculated. These data were then analyzed using pair-wise *t* tests, comparing accuracy in the sound-present and sound-absent conditions. Additionally, for experiment 1b, we performed signal detection analyses (Green and Swets 1966) using the sensitivity index ( $d'$ ) and the likelihood ratio ( $\beta$ ). A *hit* was defined as a correctly reported answer that a motion change did take place. A *false alarm* was defined as a reported motion change when there was none in the trial, a *miss* was defined as an actual motion change that was not reported and a *correct rejection* was defined as a reported “no change” when there in fact was no motion change. Paired-samples *t* tests were used for comparison of the  $d'$ - and  $\beta$ -values across the two conditions. All *t* tests yielding *p* values of 0.05 or smaller were considered to be statistically significant. Unless indicated otherwise, all statistical tests were one-tailed.

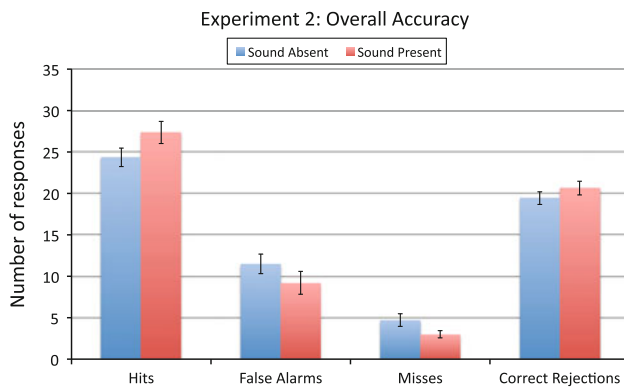
## Results

Figure 1 summarizes the main results of experiment 1a. In the sound-present condition, participants were able to detect the motion direction change (mean accuracy 79%) among on average 7.7 objects. In the sound-absent condition, they detected the motion direction change (mean accuracy 78%) among on average 6.0 objects in the sound-absent condition. The average number of displayed objects was significantly higher in the sound-present condition than in the sound-absent condition ( $t_{13} = 4.58$ ,  $P < 0.0005$ ). The difference in accuracy was short of significance ( $t_{13} = 1.77$ ;  $P < 0.1$ ; two-tailed).

The main results of experiment 1b are summarized in Fig. 2. On average, 6.9 objects were presented during a trial. Although this number varied across participants



**Fig. 1** Results of experiment 1a. Shown here is the mean number of objects presented on-screen in the sound-absent and sound-present conditions. The mean accuracy across these conditions is shown as well



**Fig. 2** Results of experiment 1b. Shown here is the average number of hits, misses, false alarms, and correct rejections. The number of hits and correct rejections is higher in the sound-present condition, compared to the sound-absent condition, while the number of misses and false alarms is lower in the sound-present condition. Results are based on a total number of 30 trials per cell

(range 4.5–11.1), it was kept constant for each participant and did not vary between the sound-absent and sound-present condition. Analyses revealed a mean accuracy of 80% in the sound-present condition and a mean accuracy of 73% in the sound-absent condition, with this difference reaching significance ( $t_{13} = 3.11$ ,  $P < 0.005$ ).

#### Sensitivity measures

The mean  $d'$  in the sound-present condition was 1.97, whereas it had a mean value of 1.42 in the sound-absent condition. This difference was significant, with  $t_{13} = 3.71$ ,  $P < 0.005$ . Computing  $\beta$ , the analysis revealed a mean value of 1.82 in the sound-present condition and a mean value of 1.58 in the sound-absent condition. These  $\beta$  estimates did not differ significantly between conditions ( $t_{13} = 0.82$ ,  $P < 0.21$ ).

#### Discussion

The present study examined the influence of an auditory, spatially uninformative stimulus on the detection of a visual motion change. This was done by determining how well participants were able to detect changes of motion direction of a single visual item among a field of continually moving objects. We manipulated the number of concurrent visual objects, by keeping the average accuracy constant (experiment 1a). Additionally, by keeping the number of objects constant, we investigated response accuracy, more specifically one's sensitivity and the likelihood ratio ( $\beta$ ) for detecting visual changes (experiment 1b).

The major finding of the current study is that the presentation of a short, uninformative sound allows

participants to better detect motion direction changes in a complex environment. As such, our results are consistent with earlier studies showing that sound can increase one's detection thresholds for visual stimuli (Stein et al. 1996; Frassinetti et al. 2002). Furthermore, the present study supports the hypothesis that a general selection mechanism, including motion selection, underlies such a sound-induced increase in visual sensitivity.

It is interesting to note that in both conditions in experiment 1a, the average number of objects was somewhat higher than the number of objects that participants are typically able to track by means of top-down attention. Although the exact number of objects that participants are able to track has been shown to vary somewhat depending on parameters such as the speed of the moving object (Alvarez and Franconeri 2007), it is generally noted that one's ability for tracking multiple objects is limited to approximately four items (e.g., Cowan 2001). Our results therefore indicate that the motion direction change may—at least to some degree—capture one's attention (see e.g., Howard and Holcombe 2010). This observation is also consistent with Girelli and Luck (1997), who found that only motion singletons automatically captured attention. Importantly, however, the results of experiment 1a also show that the presence of a spatially uninformative sound can increase the detectability of a unique stimulus among a higher number of distractors, compared to a situation where no sound is present. This observation supports a prior hypothesis that multisensory integration processes can affect the competition among multiple concurrently presented visual stimuli, by making a relevant item more salient (Van der Burg et al. 2011; Talsma et al. 2010). It should be noted that the fact that sound can affect the attentional selection of moving stimuli has some interesting implications for our understanding of how attentional selection works. When all stimuli remain at fixed locations, the relative saliency of a visual stimulus can be computed by determining the saliency of each constituent feature of a given stimulus, such as color and/or brightness and superimposing the relative contributions of each of these features onto an overall saliency map (Itti and Koch 2001). One implication of such a map is that it represents the more or less instantaneous saliency and therefore that the corresponding saliency map needs continuous updating when visual objects are continuously changing positions. Considering that, there is approximately a 240 ms time window between the onset of multisensory integration processes and the shift of attention to the location of the visual stimulus triggering the integration (Van der Burg et al. 2011), we assume that the location where attention would orient to would be slightly lagging behind the actual location of the stimulus by the time attention has shifted to this new location. In the present study, the stimulus would

have moved a little over half a degree (more precisely  $0.52^\circ$ ) during the time required to execute a shift of attention. This would presumably still be close enough to allow the visual stimulus to fall inside the attended area. If stimuli moved outside the area where attention is exogenously drawn toward, the sound-induced benefits of detecting motion direction changes might be suppressed. Thus, we would predict that the beneficial effect of the auditory stimulus no longer occurs if the speed of the visual stimuli is relatively high.

With regard to the mechanisms involved in the current study, it is interesting to notice that the sound started at the moment when the visual stimulus had already initiated moving in a new direction. The fact that the detection of a change in motion direction could be increased by such a sound makes it unlikely that the sound simply acted as a warning signal. Although currently we cannot fully rule out that a general alerting effect of the sound has contributed to our findings, we have two arguments against such an explanation. Firstly, it has been reported that increases in alertness result in a shortening of response times, at the cost of a reduction in accuracy, suggesting that highly alerted participants processed stimuli on the basis of relatively little information (Fernandez-Duque and Posner 1997; Posner 1978). The latter observation is incompatible with the current study, where we show that accuracy was increased in the sound-present condition of experiment 1b. If the tones had elicited a general alerting effect, we would have expected to find a significant change in the likelihood ratio ( $\beta$ ), as opposed to the shift in the sensitivity index ( $d'$ ) that we actually found. Secondly, alerting signals are typically most effective when presented at least 100–300 ms before the event of interest (Bertelson 1967; Niemi and Näätänen 1981; Posner and Boies 1971; McDonald et al. 2000; Spence and Driver 1997), while in the current study the sound was presented immediately after the visual event. It should be noted that the temporal alignment of the visual and auditory stimuli in our current study is compatible with the audiovisual SOA reported by Van der Burg et al. (2008, Fig. 6) to yield the strongest multisensory integration effects. Vroomen and de Gelder (2000) used a similar approach to rule out the possibility that alerting, as opposed to a perceptual process, contributed to the freezing effect reported in their study.

It should be noted that the observed audiovisual improvement in detection motion changes occurs without an explicit attentional manipulation by means of an instruction. Because our participants did not know in advance where the motion direction change would occur, our results are consistent with an already established conclusion that these increased detection rates are due to a pre-attentive process (Van der Burg et al. 2008; Driver 1996; Vroomen and De Gelder 2000). Although we note the

consistency between our findings and those of the aforementioned studies, it should be mentioned that attention was not directly manipulated in the current study. Therefore, it remains an open question to what degree the current results would have been affected by the deliberate direction of attention to specific locations on the visual field. Other studies, (Talsma and Woldorff 2005; Alsius et al. 2005; Alsius et al. 2007; Fairhall and Macaluso 2009) have shown that selective attention can influence multisensory integration or that it might even be considered a requirement for multisensory integration (Talsma et al. 2007). Presumably, several factors, including perceptual load, as well as the level of competition of stimuli among each other, may be involved in determining how attention and multisensory integration interact with each other (Talsma et al. 2010). It appears that the auditory stimulus needs to be a salient and abrupt event (as in Van der Burg et al. 2008 as well as in the current study; 2011) in order for it to be uniquely bound to a visual event (see Van der Burg et al. 2010 for more details). This also explains why Alsius and Soto-Faraco (2011) as well as Fujisaki et al. (2006) did not find support for auditory-driven visual search even though a single auditory signal was synchronized with the visual target. Based on this idea, it would follow that sounds that were part of an ongoing event or a melody (as in Vroomen and De Gelder 2000) would to a much lesser degree be able to increase the detectability of visual motion changes. Future studies should address this issue.

## Summary and conclusions

The present study investigated whether spatially uninformative sounds could influence the detection of motion-related changes in moving visual stimuli. Using a visual multiple object tracking task involving moving stimuli, we found that sounds increased one's capacity to detect visual movement changes. This result extends previous studies that have demonstrated the existence of similar auditory-induced enhancement effects for the detection of color and orientation changes. We suggest that a common attentional mechanism, subserving the selection of color, shape, and motion signals appears to be connected to multisensory integration mechanisms.

**Acknowledgments** We thank two anonymous reviewers for their constructive comments: Manon Mulckhuysen for technical assistance, as well as Erik van der Burg and Elger Abrahamse for helpful discussions. The effort for this work was supported by funding from the Institute for Behavioral Research at the University of Twente.

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

## References

- Alais D, van Boxtel JJ, Parker A, van Ee R (2010) Attending to auditory signals slows visual alternations in binocular rivalry. *Vision Res* 50(10):929–935
- Alsius A, Soto-Faraco S (2011) Searching for audiovisual correspondence in multiple speaker scenarios. *Exp Brain Res*. doi: [10.1007/s00221-011-2624-0](https://doi.org/10.1007/s00221-011-2624-0)
- Alsius A, Navarra J, Campbell R, Soto-Faraco S (2005) Audiovisual integration of speech falters under high attention demands. *Curr Biol* 15(9):839–843
- Alsius A, Navarra J, Soto-Faraco S (2007) Attention to touch weakens audiovisual speech integration. *Exp Brain Res* 183(3):399–404
- Alvarez GA, Franconeri SL (2007) How many objects can you track?: evidence for a resource-limited attentive tracking mechanism. *J Vis* 7(13):1–10
- Arrighi R, Marini F, Burr D (2009) Meaningful auditory information enhances perception of visual biological motion. *J Vis* 9(4):1–7
- Bertelson P (1967) The time-course of preparation. *Q J Exp Psychol* 19:272–279
- Bertelson P, Vroomen J, De Gelder B, Driver J (2000) The ventriloquist effect does not depend on the direction of deliberate visual attention. *Percept Psychophys* 62(2):321–332
- Bremmer F, Schlack A, Shah NJ, Zafiris O, Kubischik M, Hoffmann K, Zilles K, Fink GR (2001) Polymodal motion processing in posterior parietal and premotor cortex: a human fmri study strongly implies equivalencies between humans and monkeys. *Neuron* 29(1):287–296
- Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci* 12(12):4745–4765
- Burr D, Alais D (2006) Combining visual and auditory information. *Prog Brain Res* 155:243–258
- Calabro FJ, Soto-Faraco S, Vaina LM (2011) Acoustic facilitation of object movement detection during self-motion. *Proc R Soc B Biol Sci*. doi:[10.1098/rspb.2010.2757](https://doi.org/10.1098/rspb.2010.2757)
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10(11):649–657
- Cowan N (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci* 24(1):87
- Driver J (1996) Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature* 381(6577):66–68
- Fairhall SL, Macaluso E (2009) Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *Eur J Neurosci* 29(6):1247–1257
- Fernandez-Duque D, Posner MI (1997) Relating the mechanisms of orienting and alerting. *Neuropsychologia* 35(4):477–486
- Frassinetti F, Bolognini N, Làdavas E (2002) Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp Brain Res* 147(3):332–343
- Fujisaki W, Koene A, Arnold D, Johnston A, Nishida S (2006) Visual search for a target changing in synchrony with an auditory signal. *Proc R Soc B Biol Sci* 273(1588):865–874
- Girelli M, Luck SJ (1997) Are the same attentional mechanisms used to detect visual search targets defined, by color, orientation, and motion? *J Cogn Neurosci* 9(2):238–253
- Goodale MA, Milner AD (1992) Separate visual pathways of perception and action. *Trends Neurosci* 15(1):20–25
- Graziano MSA (2001) A system of multimodal areas in the primate brain. *Neuron* 29(1):4–6
- Green DM, Swets JA (1966) Signal detection theory and psychophysics. Wiley, New York
- Howard CJ, Holcombe AO (2010) Unexpected changes in direction of motion attract attention. *Atten Percept Psychophys* 72(8):2087–2095
- Itti L, Koch C (2001) Computational modelling of visual attention. *Nat Rev Neurosci* 2(3):194–203
- Kim J, Kroos C, Davis C (2010) Hearing a point-light talker: an auditory influence on a visual motion detection task. *Perception* 39:407–416
- McDonald JJ, Teder-Salejarvi WA, Hillyard SA (2000) Involuntary orienting to sound improves visual perception. *Nature* 407(6806):906–908
- Navarra J, Alsius A, Soto-Faraco S, Spence C (2010) Assessing the role of attention in the audiovisual integration of speech. *Inf Fusion* 11(1):4–11
- Ngo NK, Spence C (2010) Auditory, tactile, and multisensory cues can facilitate search for dynamic visual stimuli. *Atten Percept Perform* 72:1654–1665
- Niemi P, Näätänen R (1981) Foreperiod and simple reaction time. *Psychol Bull* 89(1):133–162
- Noesselt T, Bergmann D, Hake M, Heinze HJ, Fendrich R (2008) Sound increases the saliency of visual events. *Brain Res* 1220(C):157–163
- Posner MI (1978) Chronometric explorations of the mind. Lawrence Erlbaum, Hillsdale
- Posner MI, Boies SJ (1971) Components of attention. *Psychol Rev* 78(5):391–408
- Roseboom W, Nishida S, Fujisaki W, Arnold DH (2011) Audio-visual speech timing sensitivity is enhanced in cluttered conditions. *PLoS ONE* 6(4):e18309
- Soto-Faraco S, Alsius A (2007) Conscious access to the unisensory components of a cross-modal illusion. *Neuroreport* 18(4):347–350
- Soto-Faraco S, Navarra J, Alsius A (2004) Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. *Cognition* 92(3)
- Spence C, Driver J (1997) Audiovisual links in exogenous covert spatial orienting. *Percept Psychophys* 59(1):1–22
- Standage GP, Benevento LA (1983) The organization of connections between the pulvinar and visual area mt in the macaque monkey. *Brain Res* 262(2):288–294
- Stein BE, London N, Wilkinson LK, Price DD (1996) Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *J Cogn Neurosci* 8(6):497–506
- Stein BE, Jiang W, Stanford TR (2005) Multisensory integration in single neurons of the midbrain. In: Calvert G, Spence C, Stein BE (eds) *The handbook of multisensory processes*. MIT press, Cambridge
- Talsma D, Woldorff MG (2005) Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *J Cogn Neurosci* 17(7):1098–1114
- Talsma D, Doty TJ, Woldorff MG (2007) Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb Cortex* 17(3):679–690
- Talsma D, Senkowski D, Woldorff MG (2009) Intermodal attention affects the processing of the temporal alignment of audiovisual stimuli. *Exp Brain Res* 198(2–3):313–328
- Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between attention and multisensory integration. *Trends Cogn Sci* 14:400–410
- Van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J (2008) Pip and pop: nonspatial auditory signals improve spatial visual search. *J Exp Psychol Hum Percept Perform* 34(5):1053–1065
- Van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J (2009) Poke and pop: tactile-visual synchrony increases visual saliency. *Neurosci Lett* 450(1):60–64

- Van der Burg E, Cass J, Olivers CNL, Theeuwes J, Alais D (2010) Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS ONE* 5(5):e10664. doi:[10.1371/journal.pone.0010664](https://doi.org/10.1371/journal.pone.0010664)
- Van der Burg E, Talsma D, Olivers CNL, Hickey C, Theeuwes J (2011) Early multisensory interactions affect the competition among multiple visual objects. *Neuroimage* 55:1208–1218
- Van Ee R, Van Boxtel JJA, Parker AL, Alais D (2009) Multisensory congruency as a mechanism for attentional control over perceptual selection. *J Neurosci* 29(37):11641–11649
- Vroomen J, De Gelder B (2000) Sound enhances visual perception: cross-modal effects of auditory organization on vision. *J Exp Psychol Hum Percept Perform* 26(5):1583–1590
- Vroomen J, Bertelson P, De Gelder B (2001a) Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychol (Amst)* 108(1):21–33
- Vroomen J, Bertelson P, De Gelder B (2001b) The ventriloquist effect does not depend on the direction of automatic visual attention. *Percept Psychophys* 63(4):651–659