JOURNAL OF COGNITIVE PSYCHOLOGY
2021, VOL. 33, NO. 3, 282–303
https://doi.org/10.1080/20445911.2021.1896528

Audiovisual looming signals are not always prioritised: evidence from exogenous, endogenous and sustained attention

Hanne Huygelier [a], Raymond van Ee [a,b], Armien Lanssens [a], Johan Wagemans [a] and Céline R. Gillebert [a]

[a]Department of Brain and Cognition, KU Leuven, Leuven, Belgium; [b]Donders Institute for Brain, Cognition and Behavior, Radboud University, Nijmegen, Netherlands

ABSTRACT
The majority of studies showing multisensory attention benefits have focused on brief audiovisual events. Here we examined whether attention benefits can occur for auditory and visual signals that change in synchrony over time. In the first two experiments, we found no evidence that attention was captured more by synchronous compared to asynchronous audiovisual and visual looming signals. The results of our third experiment suggested that attention was not preferentially oriented towards synchronous compared to asynchronous audiovisual and visual looming signals. In the fourth experiment, we found no evidence for better sustained attention for synchronous compared to asynchronous audiovisual looming and unisensory signals. Together, these findings indicate that synchronous multisensory looming signals are not always prioritised by our information processing system. Future multisensory research should focus on more conceptual clarity and deepen our understanding of the specific stimulus, task and contextual features affecting the strength and quality of multisensory integration.

ARTICLE HISTORY
Received 6 April 2020
Accepted 23 February 2021

KEYWORDS
Multisensory integration; looming; exogenous; endogenous

Introduction

In our daily life, we continuously combine information from multiple sensory modalities to make sense of the environment around us. For instance, when a car drives towards us, we can estimate the moment it will pass by. To make such an estimation, we can flexibly switch between the visual and the auditory modality (Gordon & Rosenblum, 2005). Consequently, understanding how multisensory information guides attention is important (Ghazanfar & Schroeder, 2006). Indeed, a lot of research has been dedicated to understand how multisensory information guides the selection of information from our environment. Overall, these studies suggest that multisensory signals are prioritised, as they convey information about common sources in the environment, making multisensory information behaviourally relevant (Körding et al., 2007; Rohe & Noppeney, 2015).

Brief multisensory signals capture attention

Prioritisation of multisensory signals has been demonstrated to be stimulus-driven (i.e. exogenous) by several attention capture studies, showing that multisensory signals captured attention even when this would not be advantageous to the task goals (Matusz & Eimer, 2011; Santangelo & Spence, 2007, 2008; Van der Burg et al., 2008a, 2008b, 2009, 2011). For instance, Van der Burg et al. (2008a) used a visual search paradigm in which participants had to search for a vertical or horizontal line segment among oblique line segments. The line segments changed colour continuously in a pseudo-random fashion. If the target changed colour, it was the only line segment that did so. When the colour change of the target coincided with a tone pip, visual search was faster than in the tone-absent trials. Importantly, the search benefit was larger when the tone pip was presented 25–50 ms after the colour change of the target than

CONTACT Céline R. Gillebert celine.gillebert@kuleuven.be Department of Brain and Cognition, KU Leuven, Tiensestraat 102, box 3711, Leuven 3000, Belgium
Supplemental data for this article can be accessed at https://doi.org/10.1080/20445911.2021.1896528.

when the tone pip was presented before the colour change. These findings support the idea that the tone pip is integrated with the colour change, increasing the salience of the target in the visual search display. Furthermore, visual search for the target was faster for tone-present trials even when the tone pip occurred simultaneously with distractor colour changes on 80% of trials. The latter shows that audiovisual integration even takes place when the tone pip is completely uninformative for the visual search performance and would best be ignored.

Attention capture by brief multisensory signals has been corroborated using different paradigms, such as a simultaneity judgement paradigm (Van der Burg et al., 2008b). The simultaneity judgement paradigm is based on the principle that attended events are perceived earlier in time than unattended events (Shore et al., 2001; Stelmach & Herdman, 1991). In this paradigm, in half of the trials, two dots are presented quickly after each other and in the other half the dots are presented simultaneously. Observers are asked to respond whether the two dots appeared simultaneously or not. In a study by Van der Burg et al. (2008b), participants were more likely to perceive the two dots as having appeared at different timepoints, when the first dot was presented in close spatial proximity to a brief audiovisual signal than to a brief visual signal (Van der Burg et al., 2008b), again indicating that audiovisual signals are prioritised over visual signals.

Beyond brief audiovisual signals

So far, the majority of multisensory research has focused on a limited set of multisensory signals, among which brief audiovisual signals have received most attention. Nonetheless, continuous repetitive (i.e. synchronous) multisensory signals are also behaviourally relevant. Indeed, it would be beneficial to integrate and attend auditory and visual signals that are synchronous over time, as they likely arise from a common source in the environment (Fujisaki et al., 2006). It would also be beneficial to integrate repetitive signals with the same frequency only. As such, one can avoid that repetitive auditory and visual signals arising from a different source are erroneously integrated (Fujisaki et al., 2006). Multisensory looming signals that are synchronous over time may be particularly relevant to daily life behaviour, as looming (i.e. approaching) signals are important to attend in order to avoid collisions. Therefore it is important to know to what extent results found with brief audiovisual signals generalise to other types of multisensory signals.

Auditory looming signals affect attention for visual looming signals

There is evidence that auditory looming signals affect attention for visual looming signals. For instance, when presented with one receding and one looming visual signal paired with either a looming or receding sound, Rhesus monkeys preferentially gazed at the visual looming signal when paired with the auditory looming signal but not when paired with the auditory receding signal, suggesting a spontaneous preference for audiovisual looming signals (Maier et al., 2004). Similar findings were reported in 5-month old human infants (Walker-Andrews & Lennon, 1985).

Although it seems likely that multisensory looming integration would be frequency-specific (Fujisaki et al., 2006) and therefore facilitating effects of multisensory looming signals would also be frequency-specific, there are only two studies available, to our knowledge, that contrasted frequency-matching (i.e. synchronous) and frequency-mismatching (i.e. asynchronous) multisensory looming signals (Moors et al., 2014; van Ee et al., 2009). Both studies suppressed the visual signals from awareness, either by means of binocular rivalry or continuous flash suppression. In the study of van Ee et al. (2009), attention for the auditory signal was required to find an effect of a synchronous auditory looming signal on rivalry between a looming and rotating visual signal, while in the study of Moors et al. (2014), synchronous auditory looming signals did not affect perceptual thresholds for visual looming signals suppressed from awareness by continuous flash suppression. Thus, it is unclear to what extent conscious access to visual looming signals is required for audiovisual looming integration. In both of these studies, the visual looming signals were presented sub-threshold, which could explain why synchronous and asynchronous multisensory looming conditions did not differ or why attention was required to find multisensory benefits (Macaluso et al., 2016). That is, in the context of sub-threshold visual signals, the visual signals may not be sufficiently processed in order for them to be integrated with the auditory modality.

Thus, although studies have found positive effects of auditory looming on attention for visual looming signals, previous studies have not contrasted supra-threshold synchronous and asynchronous looming signals. Therefore, it remains unclear to what extent supra-threshold synchronous multisensory looming signals would be prioritised over supra-threshold asynchronous multisensory looming signals.

Is multisensory attention capture specific to brief signals?

Some researchers have suggested that only brief audiovisual signals can capture spatial attention. Van der Burg et al. (2010) used a visual search display in which oblique distractor lines and a horizontal or vertical target line were surrounded by an annulus. The luminance of each of the annuli was either changing according to a sine wave or square wave. In half of the trials, participants were also presented with a sound with the same temporal profile. Visual search was faster in the condition that the luminance of the annuli were changing according to a square wave paired with the square wave auditory signal compared to the condition without the auditory signal. However, the sine-wave auditory signal had no effect on visual search. The latter suggests that a sudden visual change is necessary for audiovisual signals to capture attention.

Although a sudden visual change may be necessary to exogenously capture attention, sine-wave signals may also have positive effects on attention. Marchant et al. (2012) presented a train of checkerboard patterns and pure tones either following a sinusoidal predictable timing or an unpredictable erratic timing. Participants had to detect higher contrast checkerboards or louder tones in the train of stimuli. They found that participants detected more visual and auditory targets when the auditory and visual train of stimuli was synchronous versus when the auditory and visual train of stimuli was asynchronous. The latter result suggests that synchronous stimuli may boost our ability to maintain attention over time for a repetitive task (i.e. sustained attention).

The current study

Thus, it remains unclear whether suprathreshold task-irrelevant auditory looming signals affect the attentional priority of visual looming signals in a similar fashion as reported for brief auditory signals (Matusz & Eimer, 2011; Santangelo & Spence, 2007, 2008; Van der Burg et al., 2008b, 2011). Similarly to brief audiovisual signals, audiovisual looming signals that are synchronous over time may be behaviourally relevant signals that are preferentially attended, as they likely arise from a common source in the environment (Fujisaki et al., 2006; Körding et al., 2007; Rohe & Noppeney, 2015). Moreover, it is likely that observers would only integrate multisensory looming signals matching in frequency, to avoid integrating repetitive auditory and visual signals that do not arise from a common source (Fujisaki et al., 2006). Given the behavioural relevance of synchronous audiovisual signals, we expected a preference to attend synchronous over asynchronous audiovisual looming signals, even when the synchronous audiovisual looming signal is task-irrelevant. Thus, our research question was: do human observers preferentially attend supra-threshold synchronous over asynchronous audiovisual looming signals, in a situation in which this attention bias towards synchronous multisensory signals is not relevant to the current task at hand?

In the current study, we conducted four behavioural experiments. Experiments 1 and 2 were designed to assess our original research question and thus examined exogenous spatial attention. In Experiment 1, we presented observers two supra-threshold visual looming disks without sound, or at a frequency that either matched (i.e. synchronous) or mismatched the frequency (i.e. asynchronous) of a looming sound. The luminance of one of the two visual looming signals briefly changed and participants were instructed to detect this as fast as possible. We predicted better and faster target detection when the luminance change would occur on the synchronous compared to the asynchronous audiovisual looming disk. In Experiment 2 we re-examined our original research question, using a simultaneity judgement task previously used by Van der Burg et al. (2008b). In Experiment 1 and 2 we found no evidence for differences in performance between the three audiovisual conditions.

Given the lack of exogenous spatial attention effects in Experiments 1 and 2, we then examined whether human observers would preferentially orient their attention towards supra-threshold synchronous versus asynchronous multisensory looming signals, when orienting attention towards the signals is relevant to the task at hand. Thus, we examined endogenous (i.e. voluntary) attention preferences for synchronous audiovisual looming signals in Experiment 3, but again found no differences between our audiovisual conditions.

Given the lack of spatial attention effects in our first three experiments, we evaluated a different type of attention, sustained attention, in Experiment 4. The interaction of multisensory stimuli and attention may differ for sustained versus selective attention considering differences in their neural basis. While selective attention strongly relies on a fronto-parietal dorsal attention network (Corbetta et al., 2008; Gillebert et al., 2011; Kastner & Ungerleider, 2000), other large-scale networks, such as the default-mode and ventral attention networks, seem to be involved in sustained attention (Bonnelle et al., 2011; Esterman et al., 2013; Kucyi et al., 2017; Langner & Eickhoff, 2013; Molenberghs et al., 2009). Moreover, a previous study found better detection of oddball-flashes and oddball-tones when these were presented in a synchronised sequence of auditory and visual information (Marchant et al., 2012). This effect could be mediated by sustained attention. We anticipated on finding better sustained attention for synchronous compared to asynchronous audiovisual signals. However, we again found no beneficial effect of audiovisual synchronicity. In sum, our results did not reveal beneficial effects of multisensory looming signals on attention in each of our experiments.

Experiment 1: exogenous attention

In Experiment 1, we assessed whether synchronous audiovisual looming signals were preferentially attended compared to asynchronous audiovisual and visual looming signals. To this end, we presented observers two visual looming disks without sound, or at a frequency that either matched (i.e. synchronous) or mismatched the frequency (i.e. asynchronous) of a looming sound. At pseudo-random timepoints the luminance of one of the two visual looming signals briefly changed and participants were instructed to detect this as fast as possible. Note that the auditory signals are irrelevant to the visual task, similar to previous multisensory attention capture paradigms (Matusz & Eimer, 2011; Santangelo & Spence, 2007, 2008; Van der Burg et al., 2008a, 2011). We expected a higher accuracy and faster response times for the synchronous audiovisual targets compared to asynchronous and visual looming targets without accompanying sound. In addition, we expected the difference between synchronous and asynchronous audiovisual looming targets to be most pronounced when both signals were presented simultaneously.

Method

Participants

Eighteen neurologically healthy participants took part in this study. Ages ranged from 19 to 34 years ($M = 22$, $SD = 3$). Of the 18 participants, 15 were female and 4 were left handed. All procedures in the four experiments were approved by the institute's ethical committee (SMEC G-2017 07858). Written informed consent was obtained from all participants prior to each experiment. All participants had normal or corrected-to-normal hearing and vision, and were naive to the study goals.

Apparatus

Stimuli were created and presented using PsychoPy v1.84 (Peirce, 2007) and Python 2.7. Stimuli were presented on a 24 inch gamma-corrected monitor with a resolution of 1920 by 1200 pixels and a refresh rate of 60 Hz in a dark room. All stimuli were presented on a grey uniform background at a luminance of 30 cd/m^2. Participants sat across the monitor at a distance of 54 cm with their head stabilised with a chinrest. Auditory stimuli were presented through noise cancelling headphones (Sony MDR-ZX110NA). Responses were registered with an AZERTY keyboard and participants responded with their right hand.

Stimuli

Looming signals were designed as described in Parker and Alais (2007), who found that looming signals dominated perceptual rivalry over receding signals. One looming cycle consisted of an exponential increase of the size of a disk during 80% of the period, followed by a cosine decay for 20% of the period (Figure 1(A)). Previous research has used a two-dimensional exponential increase in the size of a disk as a looming signal (Cappe et al., 2009; Dent & Humphreys, 2011; Franconeri & Simons, 2003; Maier et al., 2004, 2008; Tyll et al., 2013). In Experiments 1–3, the size change could occur at one of two frequencies: 1.25 Hz (*fast looming*) and 0.75 Hz (*slow looming*). The fast looming disk increased size from 0.75° to 2.25° over a period of 640 ms and changed back to its original size over

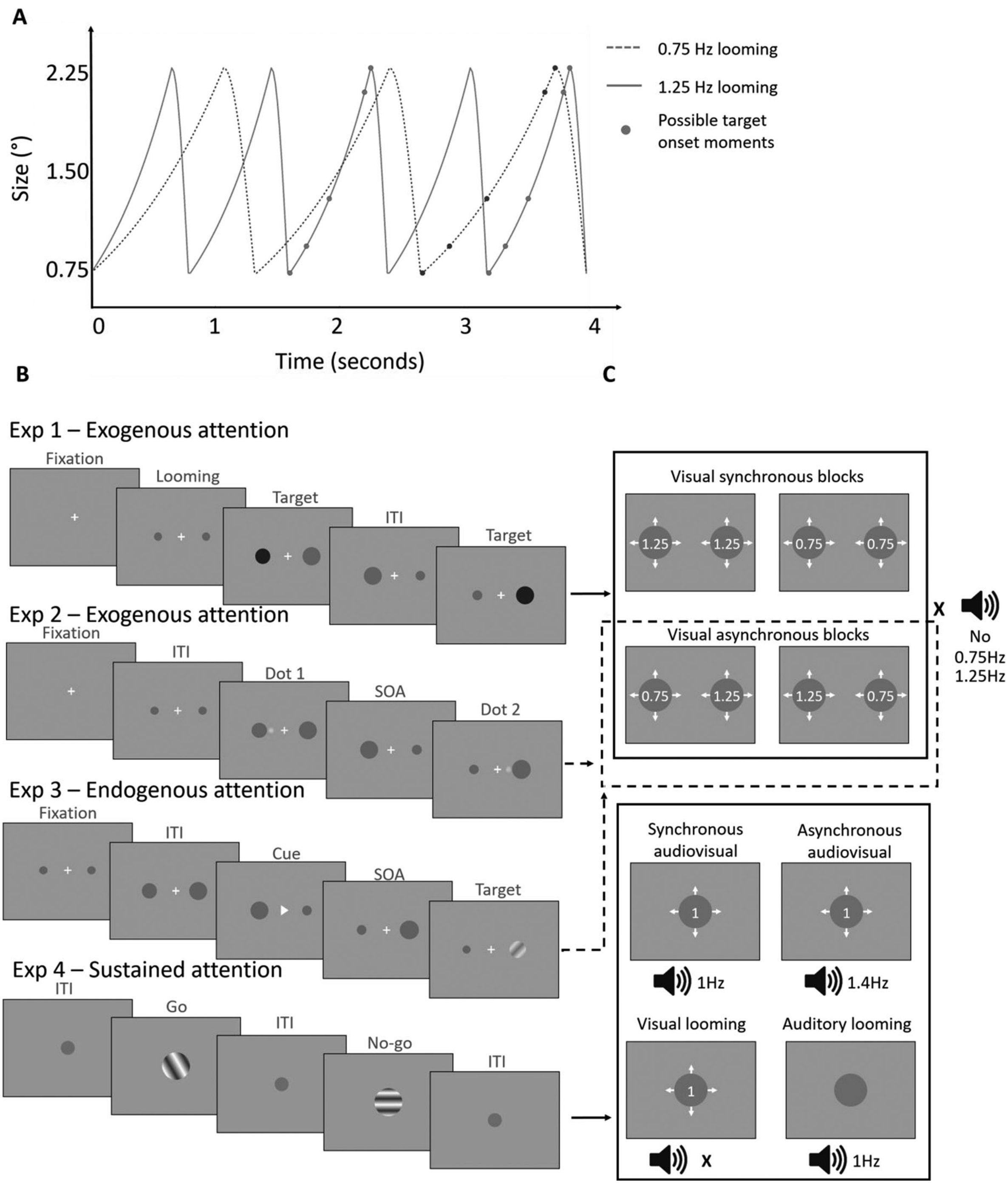

Figure 1. Overview of the design of the looming stimuli used in Experiments 1–3 (A), procedures (B) and design of experimental blocks of all four experiments (C).

a period of 160 ms for one looming cycle. The slow looming disk changed size with the same amplitude, but grew over a period of 1067 ms and changed back to its original size over a period of 267 ms for one looming cycle. The visual looming signals were continuously presented for multiple cycles. The visual looming signals were located to the right and left of a white central fixation cross of 1° at 8° eccentricity in Experiments 1–3. The two visual looming signals were presented at 10% contrast (27 cd/m^2) in Experiment 1.

A fast (1.25 Hz) and slow (0.75 Hz) auditory looming signal were created by modulating the sound pressure level of a 250 Hz pure tone using the same timing as the visual looming signals. That is, the intensity of the tone would increase from a sound pressure level of 0 dB to a sound pressure level of 52 dB exponentially and then decrease again to a sound pressure level of 0 dB following a cosine decay. Previous studies have demonstrated integration of multisensory looming signals using frequencies of 1 Hz (Maier et al., 2004, 2008; van Ee et al., 2009).

Design

We manipulated the synchronicity of the two visual looming disks, by either presenting a block of trials in which the two disks changed size at the same looming frequency (i.e. *visual synchronous blocks*) or at a different looming frequency (i.e. *visual asynchronous blocks*) (Figure 1(C)). There were two types of visual asynchronous blocks in which the spatial location of the 1.25 and 0.75 Hz looming disks (left versus right of the fixation cross) were crossed (Figure 1(C)). There were two types of visual synchronous blocks in which the looming disks either changed size at 1.25 or 0.75 Hz. Additionally, we manipulated the audiovisual synchronicity of the stimuli by pairing each of the four types of visual displays either with a 0.75, 1.25 Hz looming sound or no sound (Figure 1(C)). In the visual asynchronous blocks the looming sound would be synchronous with one of the two looming disks, while in the visual synchronous blocks the looming sound would either be synchronous or asynchronous with both looming disks. Supplementary video 1 illustrates the visual asynchronous block with the 1.25 Hz looming disk on the left side of the display paired with the 1.25 Hz auditory looming signal and supplementary video 2 illustrates the same visual asynchronous block paired with the 0.75 Hz auditory looming signal.

Procedure

The target event was a change of contrast of one of the two looming disks from 10% (27 cd/m^2) to 17.5% (24.8 cd/m^2) during 100 ms (Figure 1(B)). The target position (left or right) was determined at random but with the constraint that there were an equal number of left and right targets. The timing of target presentation was designed in such way that the size of the looming disk at the moment of target presentation would not be predictive of the likelihood of a target event and that the target could appear at a moment when the fast and slow looming disks were the same or a different size (Figure 1(A)). This was done to avoid participants attending only to specific moments of the looming sequence. Due to this procedure, the target size and inter-trial-intervals were not constant. Targets had an average size of 1.46° (SD = 0.6°, min = 0.75°, max = 2.25°) and the inter-trial-interval varied from 1 to 7 s (M = 3.78, SD = 1.06). The target size and the inter-trial-interval were independent of the main experimental manipulations. Participants were instructed to respond as quickly as possible to the target events by pressing the space bar on the keyboard.

The experimental trials were presented to participants in a blocked design with a total of 12 blocks. In 8 blocks either the fast or slow looming sound was continuously presented. In the remaining 4 blocks, the looming disks were not accompanied by a sound. Each block consisted of 44 target trials and 4 catch trials. The presentation order of the blocks was randomised across participants. The frequency of each looming disk was constant within each block. Each block lasted on average 4 min and started with the presentation of the fixation cross for 1.5 s followed by a period of 4 s in which no targets where presented. Short 1-minute breaks were mandatory in-between blocks. Before the start of the experimental blocks, 40 practice trials were administered. During the practice, feedback was provided by changing the colour of the fixation cross to green if a target was correctly detected and by means of verbal feedback given by the experimenter if targets were missed.

After the experiment, participants completed a task designed to assess whether they were able to discriminate the synchronous and asynchronous audiovisual signals (Supplementary Materials 1).

Results

We tested whether there was an interaction effect of visual and audiovisual synchronicity on accuracy using a Bayesian hierarchical logistic regression model and for response times of detected trials using a Bayesian hierarchical ex-Gaussian model. Catch trials were not included in the models. More details about the analyses are reported in Supplementary Materials 2.

There was a main effect of visual synchronicity, with lower chance to detect the target and slower responses in the visual asynchronous compared to the visual synchronous conditions (Table 1, Figure 2). In addition, there was an interaction effect of audiovisual synchronicity and visual synchronicity for detection probabilities (Table 1, Figure 2(A)), but not in the expected direction. In

Table 1. Model estimates of Experiment 1.

Predictor	Estimate	95% credible interval	
Probability to detect a target			
Intercept	0.76	0.27	1.25
Visual synchronous	0.63	0.46	0.81
No sound	0.28	0.05	0.54
Audiovisual synchronous	0.06	−0.15	0.26
Visual synchronous * No sound	−0.42	−0.67	−0.18
Visual synchronous * Audiovisual synchronous	−0.30	−0.54	−0.06
Response times for detected trials			
Intercept	0.59	0.57	0.62
Visual synchronous	−0.04	−0.05	−0.03
No sound	−0.01	−0.02	0.00
Audiovisual synchronous	0.00	−0.01	0.01
Visual synchronous * No sound	0.02	0.01	0.03
Visual synchronous * Audiovisual synchronous	0.00	−0.01	0.02

Notes: The visual asynchronous condition and audiovisual asynchronous conditions were used as the reference groups. Catch trials were not included in the models. Detection probabilities were modelled using a Bayesian hierarchical logistic regression model. Estimates are presented in log odds. Response times were modelled using a Bayesian hierarchical ex-Gaussian model. The variance of the Gaussian component and the rate of the exponential component were estimated as constant parameters independent of the conditions. More details about analyses are reported in Supplementary Materials 2.

the visual asynchronous conditions, the probability to detect a target was highest in the no-sound condition as compared to the audiovisual conditions, which did not differ from each other (Table 1, Figure 2(A)). In the visual synchronous conditions, there was no evidence for a difference in the probability to detect targets between the no-sound and asynchronous audiovisual conditions and the probability to detect targets was lower for the synchronous than for the asynchronous audiovisual conditions (Table 1, Figure 2(A)).

There was also an interaction of visual synchronicity and audiovisual synchronicity on response times, again not in the expected direction. For the visual asynchronous conditions, there were no systematic differences between the three audiovisual conditions (Table 1, Figure 2(B)). For the visual synchronous condition, responses were slower for the no-sound than the asynchronous audiovisual conditions and there was no evidence for a difference between the no-sound and the synchronous audiovisual conditions (Table 1, Figure 2(B)). The individual participant's data (Figure 2) did not reveal influential datapoints obscuring a potential attention benefit of synchronous audiovisual looming signals. Results of a discrimination task that participants completed after the experiment showed that participants were able to discriminate the synchronous and asynchronous audiovisual looming signals (Supplementary Materials 1).

Discussion

No evidence was found for faster or better contrast change detection for synchronous audiovisual

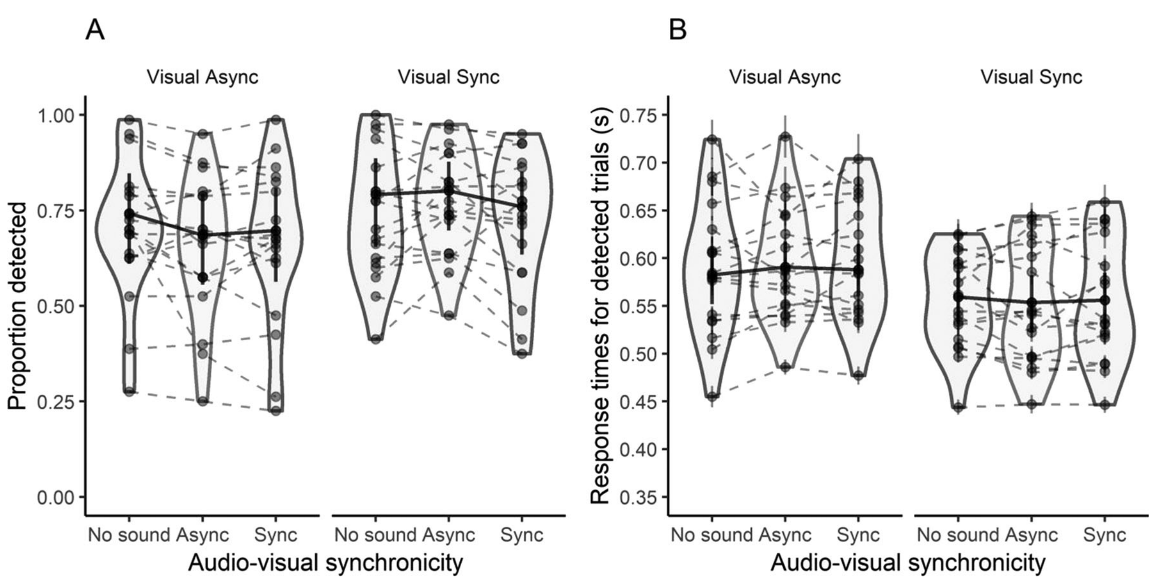

Figure 2. Results of Experiment 1. The dashed lines represent the observations of individual participants. In Panel (A) mean proportion correct is shown, while in Panel (B) median response times for detected trials and standard errors are shown for each participant. In Panel (A) the single dot and solid line represent the estimated probability of detection and in Panel (B) the estimated mean response times for detected trials. The densities represent the density of the mean proportion of detected trials (A) and of median response times for detected trials (B).

signals compared to asynchronous audiovisual and visual looming signals. The lack of an effect of audiovisual synchronicity cannot be explained by a floor or ceiling effect in performance as there was an effect of visual synchronicity on performance depending on the experimental conditions. These results suggest that synchronous audiovisual looming stimuli are not preferentially attended as compared to asynchronous audiovisual and visual looming signals. A number of studies have suggested good parallel information processing for separate hemifields (Alvarez & Cavanagh, 2005; Stoermer et al., 2014). Thus, it could be that observers were able to monitor the two locations in parallel in our experiment and did not need to prioritise one of both target locations. Moreover, although average performance of participants was not at ceiling or floor, our design may not have been optimal to measure spatial attention effects. That is, as the target size varied in Experiment 1, the saliency of target events may have varied from trial to trial. This potential mix of high- and low-salient luminance decrements may not have been ideal to measure spatial attention, as high-salient luminance decrements may create a pop-out effect and low-salient decrements may not be noticed irrespective of spatial attention. Therefore, we re-examined our research question in Experiment 2 using a simultaneity judgement task.

Experiment 2: re-examining exogenous attention with simultaneity judgements

We tested whether synchronous audiovisual signals capture attention more compared to asynchronous audiovisual signals, using a paradigm that had previously been used by Van der Burg et al. (2008b). We chose a simultaneity judgement task rather than a temporal order judgement task, as simultaneity judgments are supposed to be less sensitive to response bias than temporal order judgements (Van der Burg et al., 2008b). In line with the findings of Van der Burg et al. (2008b), we predicted that, if synchronous audiovisual signals were prioritised over asynchronous audiovisual signals, participants would be more likely to report that two dots appeared at a different timepoint when the first dot appeared in close spatial proximity to the synchronous compared to the asynchronous audiovisual and visual looming signals.

Method

Participants

A total of 15 right-handed observers participated in Experiment 2. Ages ranged from 20 to 38 years ($M = 24$, $SD = 4.7$). Of these 15 participants, 4 were men.

Stimuli

A blue dot was presented with a diameter of 0.43° at 5.6° to the left or right of the fixation cross at full contrast. As our audiovisual signal was presented at a further eccentricity (8° eccentricity) as compared to the audiovisual signal (5° eccentricity) in the experiments of Van der Burg et al. (2008b), the size of the blue dot and the distance between the dot and audiovisual looming signal were adjusted for this difference in eccentricity using the cortical magnification factor (Rovamo & Raninen, 1984). All other stimuli were exactly the same as in Experiment 1, except that the two visual looming signals were presented at 5% contrast rather than at 10% contrast.

Design

The audiovisual synchronicity of the left and right visual looming stimulus was manipulated in a blocked design (Figure 1(C)). The visual synchronous conditions were not included in the design (Figure 1 (C)), since we only predicted an effect of the synchronous audiovisual signal on the perceived timing of the two dots in the visual asynchronous conditions. In half of the trials, the time in between the first and second dot (i.e. stimulus onset asynchrony or SOA) was either 17, 33, 50 or 100 ms. Each SOA was presented an equal number of times. In the other half of the trials, the two dots appeared at the same time. The SOA was manipulated independently of all other experimental conditions. The first dot either appeared on the left or right side equally often. The first dot could appear when the looming disk had 1 out of 4 predefined sizes (i.e. 0.75°, 0.95°, 1.49° and 2.07°). The size of the looming disk at the moment the first dot appeared was manipulated independently of all other experimental conditions.

Procedure

Participants viewed a display with two looming disks and a fixation cross. In each trial, two blue dots were presented, one on the left side and the other on the right side of the display in-between the fixation cross and looming disks for 16.67 ms

(Figure 1(B)). The two dots were either presented simultaneously or with a short time-delay in between the dots. Participants were instructed to judge the simultaneity of the two dots, by responding whether the two dots appeared at the same time by pressing the "j" key or at a different time by pressing the "n" key on the keyboard.

Participants performed this task in 48 blocks, each block lasting approximately 2 min. In each block two visual looming stimuli were continuously presented. The looming frequency of the left and right disks remained constant within one block of trials. In 2/3 of the blocks a looming sound with a constant frequency of either 1.25 or 0.75 Hz was continuously presented. In 1/3 of the blocks no sound was presented. Each block consisted of 32 randomly sampled trials varying in SOA, location of the first dot (i.e. left or right side) and size of the looming disk at the onset of the first dot. Participants completed the experiment consisting of a total of 1536 trials in 2 sessions with an average of 2 days in-between sessions ($SD = 1.8$, Range: 1–7). Each session was made up of 24 blocks using the same composition of trials. The presentation order of the blocks was randomised. Before the experimental blocks, 24 practice trials were administered. During the practice trials, feedback was provided by presenting a high-pitch pure tone of 400 Hz for 250 ms for a correct response and a low-pitch pure tone of 200 Hz for 250 ms for an incorrect response.

Results

Simultaneity judgments were modelled as a function of the interaction effect of the location of the first dot (i.e. first dot on the synchronous, asynchronous audiovisual or the visual looming side) and the SOA using a Bayesian hierarchical logistic regression model. More details about the analyses are reported in Supplementary Materials 2. There was an effect of SOA on simultaneity judgements (Table 2, Figure 3 (A)). Participants were less likely to report that the two dots appeared simultaneously if there was more time in-between the two dot presentations (Figure 3(A)). The regression model suggested a small interaction effect of SOA and audiovisual synchronicity as the 95% credible interval only just excluded zero (Table 2), but this small interaction effect was not in the predicted direction (Table 2, Figure 3(A)). According to the regression model, the effect of SOA on simultaneity judgements was more pronounced for the synchronous compared to the asynchronous audiovisual signal and there was no evidence for a difference between the asynchronous audiovisual and visual looming signals (Table 2, Figure 3(A)). The estimated differences in slopes between the three audiovisual conditions were minimal (Table 2). The individual participant's raw data (Figure 3(B)) illustrates that only 1 of our 15 participants was more sensitive to short SOAs when the first dot appeared on the synchronous compared to the asynchronous audiovisual and visual looming side.

Table 2. Model estimates of Experiment 2.

Predictor	Estimate	95% credible interval	
Intercept	2.10	1.72	2.49
SOA	−0.06	−0.06	−0.05
No sound	−0.04	−0.32	0.24
Audiovisual synchronous	0.26	0.01	0.53
SOA * No sound	0.00	−0.01	0.01
SOA * Audiovisual synchronous	−0.01	−0.01	−0.00

Notes: Simultaneity judgments were modelled using a Bayesian hierarchical logistic regression model. Estimates are presented in log odds. Trials with SOAs equal to 0 ms were not included in the model as we were only interested in the extent to which participants were biased to perceive the two dots as simultaneous. The audiovisual asynchronous condition was used as the reference group. More details about the analyses are reported in Supplementary Materials 2.

Discussion

In Experiment 2, we found no evidence that synchronous audiovisual looming signals were preferentially attended as compared to asynchronous audiovisual and visual looming signals, replicating our findings of Experiment 1. The time in-between the two dots affected simultaneity judgements similarly as in the experiment of Van der Burg et al. (2008b), suggesting a successful replication of the basic paradigm. That is, participants reported that the two dots appeared simultaneously on approximately 80% of trials at the shortest SOA and on 30% of trials at a SOA of 50 ms in the study of Van der Burg et al. (2008b), which is similar to our results. Thus, it is unlikely that the general performance level of our participants can fully explain our null findings. Our findings contrast previous studies showing exogenous attentional biases for brief task-irrelevant audiovisual signals (Matusz & Eimer, 2011; Santangelo & Spence, 2007, 2008; Van der Burg et al., 2008a, 2008b). The difference in effects on spatial attention between brief audiovisual versus looming signals may be limited to the

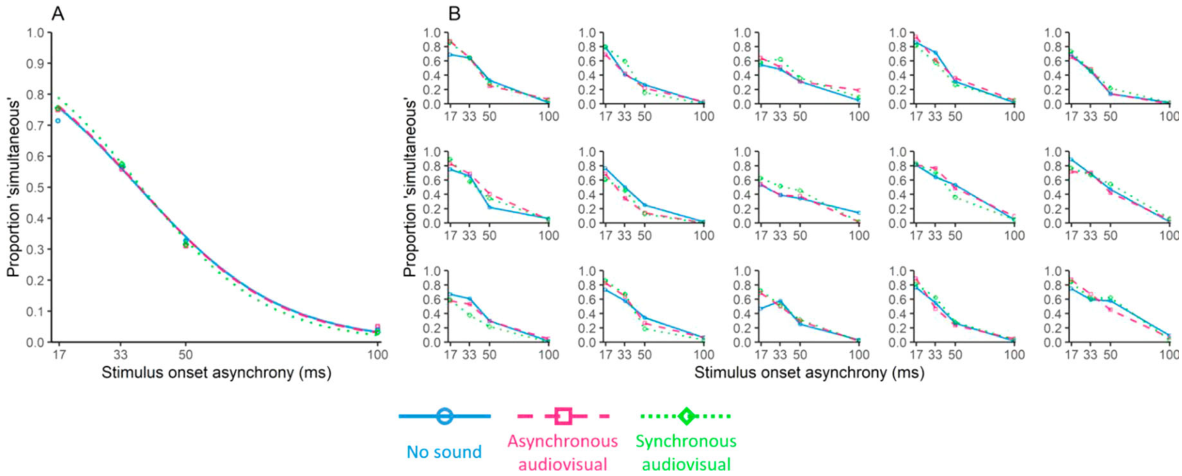

Figure 3. Results of Experiment 2. The estimated regression lines for the three audiovisual conditions are presented in panel (A) and the observed data of single participants are presented in panel (B). (To view this figure in colour, please see the online version of this journal.)

context of exogenous spatial attention. Given the fact that the audiovisual synchronicity was irrelevant to the task of luminance detection and simultaneity judgements in Experiments 1 and 2 and that audiovisual looming signals were presented continuously, the spontaneous attentional preference in favour of these signals may have been overruled by the endogenous control system (Bacon & Egeth, 1994; Posner, 1980). Therefore, our first two experiments do not rule out the possibility that synchronous audiovisual signals are preferentially attended when being more relevant for the task. Although the synchronous audiovisual looming signal did not capture attention more effectively as compared to the asynchronous audiovisual looming signal, it could be that it is easier to re-orient attention to synchronous than asynchronous audiovisual looming signals in an endogenous way. For this reason, we further tested whether there would be an endogenous attentional preference for synchronous audiovisual looming signals in Experiment 3.

Experiment 3: endogenous attention

Given the lack of prioritisation of the synchronous audiovisual looming signal in Experiments 1 and 2, we assessed whether top-down orientation of attention is differently modulated by the synchronicity of the audiovisual looming signal in Experiment 3. We presented a central arrow cue that predicted the target location in the majority of trials. In the trials in which the cue did not validly indicate the target location, covert attention was expected to be re-oriented from the cued towards the target location (Posner, 1980). We predicted faster response times and more accurate responses to targets when attention was re-oriented from an asynchronous towards a synchronous audiovisual looming signal than when attention was re-oriented from a synchronous towards an asynchronous audiovisual looming signal. In addition, we expected faster response times and more accurate responses when attention was re-oriented from an asynchronous towards synchronous audiovisual looming signal compared to when attention was re-oriented from a visual looming to another visual looming signal. Effects for this Experiment were expected to be most pronounced for response times, in accordance with previous studies using this paradigm (e.g. Bartolomeo & Chokron, 2002; Losier & Klein, 2001; Posner, 1980; Rengachary et al., 2009).

Additionally, in Experiment 3, we asked participants to perform a multisensory looming synchronicity detection task prior to the experiment. It has previously been shown that arbitrary associations between pairs of visual and auditory stimuli are rapidly acquired and that these acquired multisensory associations can bias binocular rivalry (Einhäuser et al., 2017; Piazza et al., 2018). For this reason, we anticipated that prior experience with the synchronous and asynchronous multisensory looming signals could facilitate integration of the signals during the cueing task, enhancing the chance that the synchronous multisensory looming signals are prioritised.

Method

Participants

A total of 25 healthy observers participated in Experiment 3. Ages ranged from 19 to 24 years ($M = 21$, $SD = 2$). There were 8 men and 1 left-handed participant.

Stimuli

Gaze position was monitored with an SMI Red eye tracker (Supplementary Materials 3). The target stimulus was a sinewave grating of 45° or 135° orientation of a spatial frequency of 2°. The cue was an equilateral white triangle oriented either to the left or right with a size of 0.75°. All other stimuli were exactly the same as in Experiment 1, except that the two visual looming signals were presented at 5% contrast.

Design

The experiment consisted of the visual asynchronous blocks paired either with the 1.25 or 0.75 Hz looming sound or no sound (Figure 1(C)). For each block, the endogenous cue was either oriented towards the left side or the right side with equal probability and the target appeared either on the left or right side with equal probability. The grating orientation was either 45° or 135°, equally often. The target location and grating orientation were manipulated independently. The cue direction was manipulated independently of the grating orientation. In 80% of trials the central cue pointed towards the same side as where the target appeared (valid), and in 20% of trials the cue pointed in the opposite direction as where the target appeared (invalid).

Procedure

Six experimental blocks of 100 trials each were presented. Each block consisted of one visual display paired with one looming sound condition. The order of the blocks was randomly determined per participant. In each block of 100 trials, 80 trials consisted of a valid cued target and 20 of an invalid cued target. At the start of each block, a fixation cross was presented for 1.5 s followed by the two looming disks for 4 s. During each trial, a cue was presented for 200 ms, then after a cue-target interval of 200 ms, a target was presented for 200 ms. Participants were instructed to report the orientation of the target grating by pressing either the "I" or "J" key on the keyboard. Before the experiment, participants completed a task in which they had to judge whether the audiovisual looming signal was synchronous or asynchronous in order to increase attention for audiovisual synchronicity and train individuals in discriminating the two signals (Supplementary Materials 1).

Results

We analysed response times using a Bayesian hierarchical ex-Gaussian regression model and accuracy using a Bayesian hierarchical logistic regression model as a function of the interaction between cue validity and audiovisual synchronicity. More details about the analyses are reported in Supplementary Materials 2. There was a main effect of cue validity but no evidence for a main effect of audiovisual synchronicity nor evidence for an interaction effect of audiovisual synchronicity and cue validity on accuracy (Table 3, Figure 4(A)). Orientation discrimination was more accurate in the validly cued than invalidly cued trials and there was no evidence for an effect of audiovisual synchronicity (Figure 4(A)).

There was no evidence for an interaction of audiovisual synchronicity and cue validity in the predicted direction for response times (Table 3, Figure 4(B)). There was a main effect of cue validity

Table 3. Model estimates of Experiment 3.

Predictor	Estimate	95% credible interval	
Probability of accurate discrimination			
Intercept	2.71	2.29	3.16
Valid cue	0.55	0.16	0.93
No sound	0.06	−0.27	0.42
Audiovisual synchronous	−0.27	−0.58	0.05
Valid cue * No sound	0.14	−0.28	0.54
Valid cue * Audiovisual synchronous	0.25	−0.13	0.63
Response times for accurate discriminations			
Intercept	783	738	830
Valid cue	−103	−141	−65
No sound	16	−10	43
Audiovisual synchronous	18	2	34
Valid cue * No sound	−19	−43	4
Valid cue * Audiovisual synchronous	−20	−37	−3
Sigma intercept	4.2	4.1	4.3
Sigma Valid cue	−0.1	−0.2	−0.0
Tau intercept	5.5	5.4	5.5
Tau Valid cue	−0.2	−0.3	−0.2

Notes: The invalid cue and audiovisual asynchronous conditions were used as the reference group. Accuracy was modelled using a Bayesian hierarchical logistic regression model. Estimates are presented in log odds. Response times were modelled using a Bayesian hierarchical ex-Gaussian model. The variance of the Gaussian component and the rate of the exponential component were allowed to vary between the invalid and valid condition. Estimates of sigma and tau are in log scale, while the other estimates are in response times (milliseconds). More details about analyses are reported in Supplementary Materials 2.

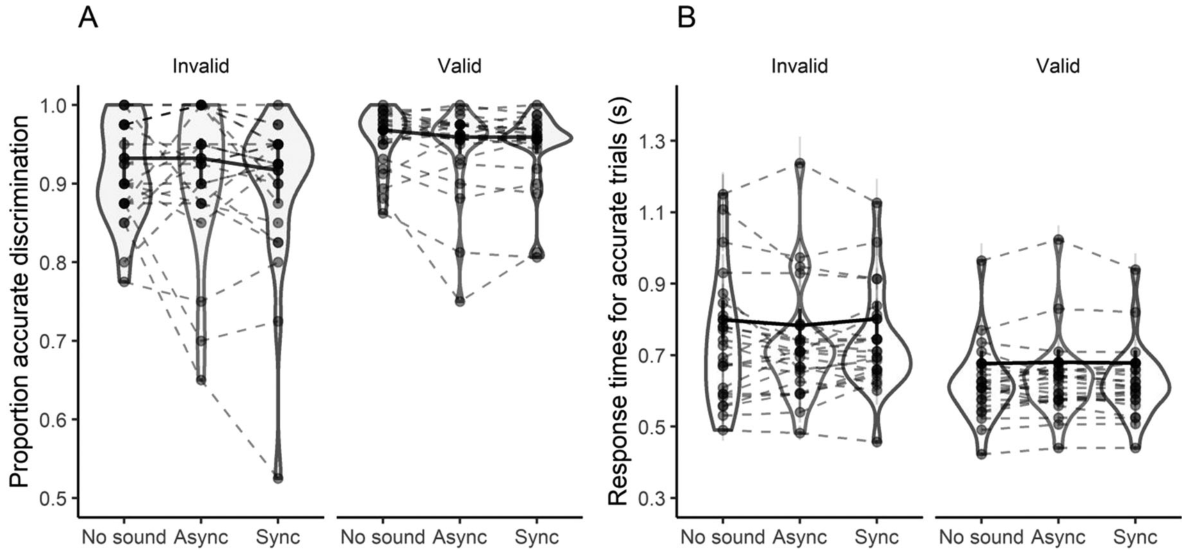

Figure 4. Results of Experiment 3. The dashed lines represent the observations of individual participants. In Panel (A) mean proportion correct is shown, while in Panel (B) median response times for accurate trials and standard errors are shown for each participant. In Panel (A) the single dot and solid line represent the estimated probability of accurate discrimination and in Panel (B) the estimated mean response times for accurate trials. The densities represent the density of the mean proportion of accurate trials (A) and of median response times for accurate trials (B).

on response times. Responses were slower in the invalidly cued compared to validly cued trials. The individual participant's data (Figure 4) do not reveal influential datapoints obscuring a potential attention benefit of synchronous audiovisual looming signals. Results taking into account eye movements led to the same conclusions (Supplementary Materials 3). The synchronicity detection task administered to participants prior to the main experiment revealed that participants were able to detect the synchronicity of the audiovisual looming signals (Supplementary Materials 1) and that there was no relation between the speed to detect the synchronicity and the effect of the audiovisual synchronicity in the main experiment (Supplementary Materials 1).

Discussion

The results of Experiment 3 did not show an endogenous attention preference in favour of synchronous audiovisual compared to asynchronous audiovisual or visual looming signals. These results cannot be fully explained by a floor or ceiling effect as there was a cue validity effect on performance and because effects were expected to be most pronounced for response times (Bartolomeo & Chokron, 2002; Losier & Klein, 2001; Posner, 1980; Rengachary et al., 2009). Thus, Experiments 1, 2 and 3 all suggest that synchronous audiovisual signals were not preferentially attended, neither exogenously, nor endogenously. In Experiment 3, we manipulated the prior experience with the audiovisual looming signals in order to increase the chance that the signals would be integrated while performing another task. Even in this context, we found no evidence for prioritisation of synchronous over asynchronous audiovisual or visual looming signals.

Experiment 4: sustained attention

In a last experiment, we addressed the effect of synchronous audiovisual signals on sustained attention. To this end, we used an adjusted version of the sustained attention to response task (SART) which is a repetitive task in which observers must monitor a stream of numbers from 1 to 9 that are presented at a fixed temporal frequency and in an unpredictable or fixed sequence (Robertson et al., 1997). The participant must respond to each number (*go trials*) except to number "3" (*no-go trials*). Using an adapted version of the SART task, we measured response time variability and responses on no-go trials (*false alarms*) as indices of sustained attention. In previous studies, researchers have used different ways to analyse response times on the SART with some extracting the slow and fast fluctuations using Fourier analyses (Johnson, Kelly, et al., 2007; Johnson, Robertson, et al., 2007) and others using ex-Gaussian analyses to extract different RT components (Tarantino et al., 2013). Previous research using ex-Gaussian analyses of response times showed the largest differences between individuals with ADHD and healthy controls in the rate of the exponential component, suggesting that this component of the response times may represent sustained attention functioning the best (Tarantino et al., 2013). We predicted less response time variability and fewer false alarms in the synchronous audiovisual condition compared to the asynchronous audiovisual and unisensory conditions.

Method

Participants

A group of 39 healthy volunteers participated in this experiment. The sample mostly consisted of right-handed individuals ($n = 33$) and females ($n = 33$). Ages ranged from 18 to 31 years ($M = 19.7$, $SD = 2.3$).

Stimuli

A disk was presented at fixation either changing size according to a looming function or at a constant size. The looming disk changed size going from a minimum of 0.5° to a maximum of 2.5°. The static disk was presented at a size of 2.5°. The disk changed size at 1 Hz and auditory looming signals changed at 1 Hz or 1.4 Hz. The auditory looming frequency was either 1 or 1.4 Hz and the visual looming frequency was 1 Hz. We adjusted the SART in such way that the target stimuli would be a feature-change of the looming disks, to avoid that participants would ignore the visual looming signal. The target stimuli were gratings created by changing the texture of the disk. The gratings had 8 possible oblique orientations ranging from 110° to 250° with 20° in-between each orientation. The target-grating had a horizontal orientation (270°). The gratings were presented for 233 ms and at a frequency of 1 Hz at the moment the visual looming disk had its maximum size.

Design

We compared sustained attention between four within-subject conditions. In the synchronous audiovisual condition the visual and auditory looming signals both had a frequency of 1 Hz (Sync), while in the asynchronous audiovisual condition the visual looming had a frequency of 1 Hz and the auditory looming had a frequency of 1.4 Hz (Async). In the auditory looming condition (AL), the visual disk was presented at a constant size and an auditory looming sound of 1 Hz was presented. In the visual looming condition (VL), the visual disk changed size at 1 Hz but no auditory looming sound was presented. These four within-subject conditions were presented using a blocked design. The order of blocks was counterbalanced between subjects with four orders: (1) Sync—Async—VL—AL, (2) Async—AL—Sync—VL, (3) AL—VL—Async—Sync and (4) VL—Sync—AL—Async. This counterbalancing scheme ensures that each condition occurs equally frequently at each time-point and that each condition precedes every other condition exactly once, controlling for carry-over effects (Maxwell & Delaney, 2003). For each of these 4 orders a minimum of 9 participants were tested. Participants were assigned to a counterbalancing order at random.

Procedure

On each trial one grating was presented. For 8 of the 9 gratings (oblique orientations) the participant was instructed to press the space bar on the keyboard (go-trials). For 1 of the 9 gratings (horizontal orientation) the participant had to withhold his response (no-go trials). The timing of the no-go trials was unpredictable and occurred at a frequency of 1 out of every 9 trials. Each block of the experiment had a total of 360 trials (i.e. 320 go trials and 40 no-go trials). Each block took 6 min to complete and there was a 4-minute break in-between each block. Before the experimental trials, participants completed 36 practice trials. During the practice trials, participants performed the task in the visual looming condition and received auditory feedback on their performance, by presenting a low-pitch pure tone of 200 Hz for 250 ms for an incorrect response.

Results

The mean response times, standard deviation and rate of the exponential component of the ex-Gaussian distribution was estimated with a Bayesian hierarchical ex-Gaussian model (Supplementary Materials 2). The estimated mean response times, standard deviations and rates are visualised in Figure 5 and model results are reported in Table 4.

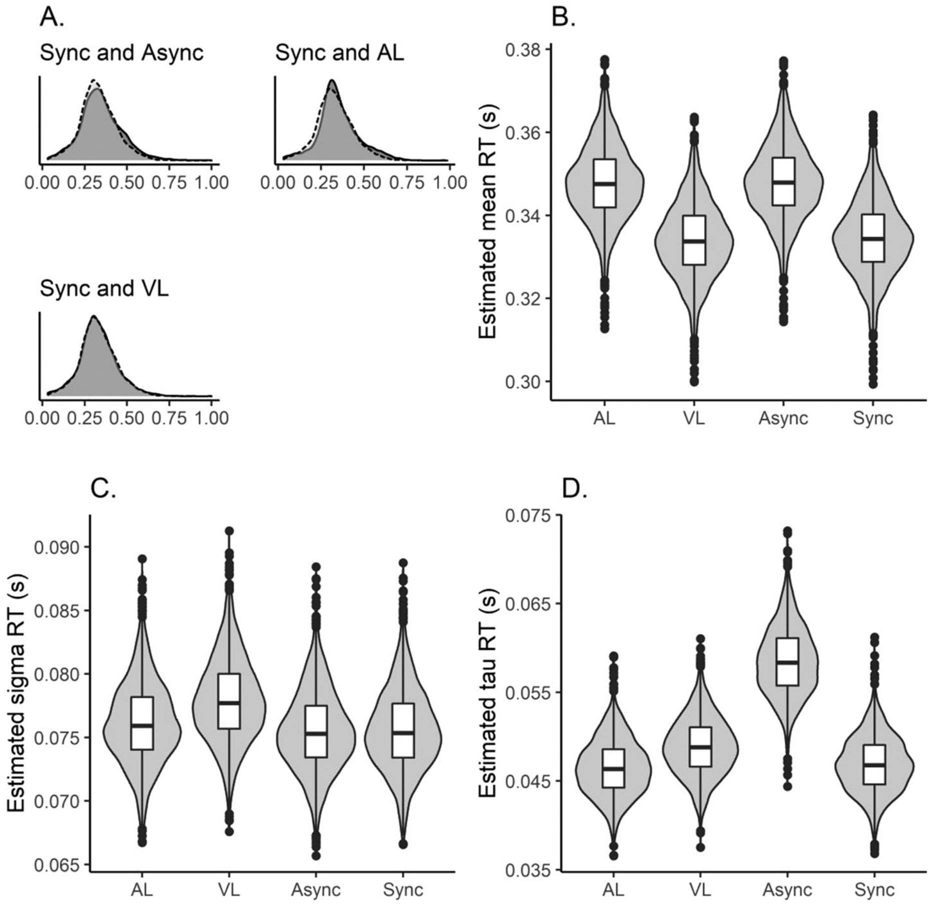

Figure 5. Results of Experiment 4. In Panel (A) observed response time distributions are shown with the synchronous condition displayed in light grey in addition to each of the other conditions in dark grey. In Panel (B) estimates of mean response times, in Panel (C) estimates of sigma and in panel (D) estimates of tau are shown.

The mean response times were lower in the synchronous compared to the asynchronous condition and auditory looming condition, but there was no evidence for a difference with the visual looming condition (Table 4, Figure 5). There was no evidence for a difference in the standard deviations of the Gaussian component between the synchronous and asynchronous condition and the synchronous and auditory looming condition. There was a difference in the standard deviation between the synchronous and visual looming condition (Table 4, Figure 5). There was a smaller exponential rate for the synchronous than asynchronous condition, but no evidence for differences between the synchronous and other conditions (Table 4, Figure 5). In sum, there was no evidence for responses that were faster or less variable in the synchronous compared to the three other conditions. In addition, a Bayesian logistic regression model revealed no evidence for any differences between conditions in false alarms (Table 4). Misses on go-trials were not analysed as they occurred too infrequently

Table 4. Results of Experiment 4.

	Ex-Gaussian						Logistic	
	Mean RT (ms)		Sigma RT (log)		Tau RT (log)		False alarms	
Predictor	*E*	95% CI	*E*	95% CI	*E*	95% CI	*E*	95% CI
Intercept	348	329, 367	4.32	4.24, 4.41	4.07	3.94, 4.19	0.14	−0.25, 0.55
Sync	−14	−16, −11	0.00	−0.03, 0.03	−0.22	−0.29, −0.15	0.09	−0.11, 0.30
AL	−0	−3, 2	0.01	−0.02, 0.04	−0.23	−0.30, −0.15	0.09	−0.18, 0.38
VL	−14	−17, −12	0.03	0.0, 0.06	−0.18	−0.26, −0.10	0.23	−0.01, 0.48

Notes: Estimates of mean response times are in ms, while estimates of sigma and tau are in log scale. The Async condition is the reference group. Estimates of false alarms are in log odds.

to derive reliable estimates (Supplementary Materials 2).

Discussion

Similar to our previous findings we did not find support for multisensory benefits. In line with our first three experiments, these data suggest that there is also no benefit of synchronous audiovisual looming signals in the context of sustained attention. Previous studies have found better detection of oddball-tones and oddball-flashes when these events occurred within a synchronised audiovisual sequence of flashes and tones, irrespective of whether the audiovisual sequence of flashes and tones was irregularly or regularly timed (Marchant et al., 2012). This suggests that synchronous audiovisual signals can enhance our ability to monitor streams of information (both predictable and unpredictable) to detect oddball-stimuli, which may partially relate to increased levels of sustained attention while monitoring the streams of information. One important difference between the study of Marchant et al. (2012) and our study is the fact that targets could appear in both sensory modalities in Marchant's study, while this was not the case in our study. Thus, together, these studies suggest that divided attention across both sensory modalities is required to find beneficial effects of synchronous multisensory signals on sustained attention. The latter is also in line with the finding that attention for the auditory modality was required to find an effect of auditory looming signals on rivalry between a rotating and looming signal (van Ee et al., 2009). Thus, the lack of a facilitating effect of synchronous multisensory signals on sustained attention in our experiment may be limited to the context in which only one sensory modality is task relevant.

General discussion

We examined whether attention benefits can occur for auditory and visual signals that change in synchrony over time. In Experiments 1 and 2, we found no evidence for an exogenous spatial attention bias favouring synchronous over asynchronous audiovisual and visual looming signals. In Experiment 3, we found no evidence for endogenous spatial attention orientation favouring synchronous over asynchronous audiovisual and visual looming signals and in Experiment 4, we found no improvement in sustained attention for synchronous compared to asynchronous audiovisual looming and unisensory signals. In sum, we found no evidence for attentional benefits of synchronous over asynchronous audiovisual looming signals and unisensory signals. These experiments indicate that not all types of multisensory events are preferentially attended. In the next sections, we discuss different potential explanations for our findings.

Can human observers integrate frequency-matched auditory and visual looming signals?

One possible explanation for our null findings is that looming signals cannot be integrated across sensory modalities irrespective of the available processing resources. However, this idea does not align with previous studies demonstrating multisensory integration of looming signals (Cappe et al., 2009; Conrad et al., 2013; Maier et al., 2004; Walker-Andrews & Lennon, 1985). All of these studies contrasted a looming auditory (i.e. increasing in amplitude) paired with a looming visual signal (i.e. increasing in size) with a condition in which a looming auditory signal is paired with a receding visual signal or vice versa. Thus, based on these studies we know that auditory and visual signals are integrated when they both increase in size or amplitude, but not when one increases and the other decreases. The studies of van Ee et al. (2009) and Moors et al. (2014) contrasted synchronous and asynchronous multisensory looming signals, while the visual looming signals were suppressed from awareness using continuous flash suppression or binocular rivalry. Thus, prior studies have not directly investigated integration of supra-threshold synchronous auditory and visual looming signals.

Nevertheless, we argue that our results cannot be fully explained by the lack of multisensory looming integration. A first reason is that performance was not better in the audiovisual compared to unisensory conditions in our four experiments. That is, if integration of looming signals was not specific to a frequency, the auditory looming signals would be integrated with both visual looming signals, and then one would still expect a multisensory benefit of the asynchronous and synchronous multisensory signals over the unisensory conditions, which was not present in our results.

A second reason is that participants had no difficulty identifying the synchronous and asynchronous audiovisual looming signals in our synchronicity discrimination and detection tasks (Supplementary Materials 1). Moreover, although there were interindividual differences in the ability to detect the audiovisual synchronicity and in the effect of the multisensory synchronous signals on attention in Experiment 3, these were not associated to each other (Supplementary Materials 1). These data show that the ability to integrate the auditory and visual looming signals did not predict the presence or absence of multisensory attention benefits.

Third, even when we first trained our participants in detecting the synchronicity and asynchronicity of the audiovisual looming signals, we still found no multisensory attention benefits in Experiment 3. The latter is noteworthy, given earlier studies showing biased binocular rivalry based on newly acquired arbitrary associations of auditory and visual signals (Einhäuser et al., 2017; Piazza et al., 2018).

Lastly, at the theoretical level, it seems unlikely that human observers would not be able to integrate information across sensory modalities matching in temporal frequency (Fujisaki et al., 2006). Multisensory integration based on matching temporal frequencies would be beneficial to daily life functioning in order to avoid that repetitive auditory signals are integrated with repetitive visual signals that do not arise from a common source (Fujisaki et al., 2006). For these reasons, we argue that it is unlikely that supra-threshold auditory and visual looming signals cannot be integrated.

Multisensory integration: an early (pre-attentive) or late (attention-modulated) process?

Our findings could also relate to the hypothesis that integration of looming signals does not occur at early information processing stages, but rather that the integration of the signals is modulated by top-down attentional resources. If attention is required for multisensory looming integration and attention is engaged with an orthogonal visual task, integration may not take place and the visual looming signal matching the task-irrelevant auditory looming signal will not be preferentially attended. This explanation for our findings aligns with studies showing that, although separate systems underlie visual and auditory perception, attentional control can have shared resources (*supra-modal attention*) for the different sensory modalities (Alais et al., 2010; Denham et al., 2018; Farah et al., 1989; Spagna et al., 2015).

Much research has been dedicated to better understand to what extent multisensory integration occurs at early (pre-attentive) versus late (attention-modulated) processing stages, with many conflicting findings. Multisensory illusions such as the beep-flash (Shams et al., 2000) and McGurk illusion (Mcgurk & Macdonald, 1976) led to a common view in the field that multisensory integration occurs at a pre-attentive early information processing stage. This idea was supported by studies, such as the one by McDonald et al. (2000) who demonstrated that task-irrelevant brief sounds increased visual detection sensitivity. However, later studies showed that the McGurk and beep-flash illusions were modulated by attentional resources (Alsius et al., 2005; Mishra et al., 2010), questioning the pre-attentiveness of audiovisual integration (Talsma et al., 2010). Moreover, it has been shown that not everyone is susceptible to the McGurk illusion (Mallick et al., 2015) and that audiovisual speech integration is susceptible to top-down attention modulations (Alsius et al., 2005), further strengthening the idea that multisensory integration may not be as pre-attentive and early as originally thought.

There have been many attempts to integrate the conflicting findings on early versus late multisensory integration. Many of the existing accounts focused on a single factor. For instance, it has been suggested that low-contrast, low spatial frequency, transient visual signals that maximally engage the magnocellular pathway are optimal signals for early multisensory integration (Jaekl et al., 2014). Others have proposed that the competition between different environmental stimuli is an important explanatory factor. Talsma et al. (2010) hypothesised that competition between multiple low-salient signals within sensory modalities may lead to a lack of early multisensory integration. However, Spence and Santangelo (2009) suggested that attentional capture by multisensory signals, which is typically viewed as an index of early integration, is more likely to occur in the context of high stimulus competition. Spence and Santangelo (2009) reasoned that unisensory cues would be as effective as multisensory cues when processing resources can be entirely dedicated to the cues, while multisensory cues will be more effective than unisensory cues when processing resources are dedicated to other stimulus information or concurrent tasks.

Other accounts have focused on the features that can be used to bind the signals across modalities. It has been suggested that audiovisual signals that co-occur in close temporal or spatial proximity are more likely to be integrated at early information processing stages than more complex signals that are synchronous over time (Macaluso et al., 2016). The spatial proximity of the signals may explain our findings. Previous studies found that task-irrelevant auditory looming signals presented simultaneously with a grating or with a cue-target onset asynchrony of 250 ms increased orientation discrimination sensitivity for the grating (Glatz & Chuang, 2019; Leo et al., 2011). Such a facilitating effect of looming sounds on orientation sensitivity was however not found in a patient with a lesion in V1 when the visual signals were presented in the blind visual field (Cecere et al., 2014). Moreover, this facilitating effect was only found for looming sounds presented on the same side of the visual stimulus (Leo et al., 2011). Thus, perhaps looming sounds only affect visual information processing at an early stage when they arise from the same spatial location as the visual signal. The looming sounds in our experiments were presented binaurally, potentially introducing a spatial ambiguity and leading to integration at lager stages of information processing.

In contrast to the many univariate explanations, some researchers have proposed that there is need for a multivariate framework to understand the interaction of attention and multisensory integration (Macaluso et al., 2016; ten Oever et al., 2016). The multivariate complexity of this problem somewhat contrasts the lack of multivariate studies on multisensory integration that have parametrically manipulated multiple stimulus-, task- and contextual factors in a single study. This practice may have led to a tendency to adhere to univariate views on multisensory integration. Therefore, future multisensory research should focus on study designs that systematically manipulate stimulus, task and contextual features that could affect integration. In addition, a recent study emphasised the importance of conceptual clarity in the field of multisensory research by showing that multisensory benefits are not the same as multisensory integration (Innes & Otto, 2019). Innes and Otto (2019) proposed that computational models may be a way forward to unify the field of multisensory research as they can offer more conceptual clarity.

Integrated, but not prioritised?

Stimulus-driven versus goal-driven attention

Finally, it is likely that the multisensory looming signals were not prioritised although the signals were integrated. Although attention capture paradigms have consistently shown that multisensory signals are preferentially processed (Santangelo & Spence, 2007, 2008; Van der Burg et al., 2008a, 2008b), there has been debate in the domain of attention research about the extent to which purely bottom-up stimulus-driven attention biases exist (Folk et al., 1992). Folk et al. (1992) suggested that attention capture effects, typically viewed as the hallmark of stimulus-driven attention, still depend on task demands. In line with his hypothesis, Folk et al. (1992) demonstrated that a task-irrelevant cue only captured attention when it shared task-relevant stimulus features with the target stimulus.

In line with Folk's hypothesis, it could be that the most important difference between our study and previous multisensory capture studies (Matusz & Eimer, 2011; Santangelo & Spence, 2007, 2008; Van der Burg et al., 2008a, 2008b, 2011) is not necessarily the type of audiovisual signal, but rather that there was a difference in the features of the targets (brief) and multisensory signals (continuous) in our experiments which resulted in a lack of attention biases towards the multisensory signals. This suggests that previous multisensory capture studies may have only found their effects because the brief nature of the audiovisual signal matched an attentional task set of participants to prioritise all brief signals. Moreover, this also raises the possibility that audiovisual looming signals would be prioritised in a situation in which participants are required to detect continuous targets. However, this account seems unlikely given the fact that even visual static signals which are typically considered as behaviourally irrelevant or weakly salient signals can capture spatial attention when participants are required to detect dynamic target events (Burnham & Neely, 2008).

Moreover, Franconeri and Simons (2003) used a visual search task in which participants were instructed to report whether a circular array of 3, 5 or 7 letters contained the target letters "U" or "E". The search array was preceded by a cue array of number 8 shapes, of which a single cue was either increasing in size (i.e. looming) or decreasing in size (i.e. receding) while the other cues had a constant size. The looming and receding cues could either appear on the same location as a target (i.e. valid) or on another location (i.e. invalid). With this paradigm, Franconeri and Simons (2003) found a significant interaction of set size and validity for the looming cues, but not for the receding cues. This suggests that the looming cues captured attention better than the receding cues. There was no reason for participants to prioritise the looming cue over the receding cue, as both cues were equally relevant to the current task of detecting either letter "U" or "E". Hence, regardless of the current goals of the participants, the looming cue captured attention more than the receding cue. These results suggest that attention can be captured even when the task-irrelevant signal (i.e. looming) does not share features with the task-relevant target (i.e. brief onsets). However, the study by Franconeri and Simons has been critiqued by Abrams and Christ (2005), suggesting that it could have been the onset of the motion that captured attention rather than the motion itself.

Habituation and mental fatigue

Another possible explanation for our null results may relate to habituation. That is, the signal may initially be prioritised, but may then quickly lose its priority due to habituation. Indeed, brief task-irrelevant visual signals lose their spatial attention capture effect over repeated trials (Turratto & Pascucci, 2016). However, our results cannot be fully explained by habituation, since we presented the synchronous and asynchronous audiovisual looming signals simultaneously for equal durations. Previous research has shown differences in prioritisation of task-irrelevant signals across long time-periods using similar designs. For instance, Parker and Alais (2007) showed that looming signals dominated binocular rivalry for longer time periods than receding signals across periods of 8 min. Moreover, they did not find evidence for a reduction in the dominance of the looming signal over time. In addition, Conrad et al. (2010) showed that an auditory signal increased the dominance duration of a random-dot kinematogram that matched the auditory signal when signals were presented for periods of approximately 3 min.

There is also evidence for persistent attention biases in other paradigms than binocular rivalry. For instance, observers were simultaneously presented with four streams of shapes in a study by Zhao et al. (2013). One of the streams consisted of a statistically regular sequence of shapes (i.e. predictable sequence), while another stream consisted of a random sequence. In the two other streams, the same shape was constantly presented. Participants had to perform a visual discrimination task on a target shape presented at pseudo-random time points in one of the streams. Attention was spontaneously biased towards the stream consisting of the statistically regular sequence of shapes. These effects were established across blocks of trials that lasted approximately 5 min (Yu & Zhao, 2015; Zhao et al., 2013). More importantly, this bias persisted during a test phase when random sequences of shapes were presented in all four streams (Yu & Zhao, 2015). Thus, certain signals seem to be prioritised over other signals even at larger timescales.

Moreover, in addition to habituation potentially affecting stimulus-driven prioritisation of audiovisual looming signals, mental fatigue could also play a role in our experiments (Boksem et al., 2005; Boksem & Tops, 2008; Faber et al., 2012). That is, previous studies have shown that task-irrelevant signals are less easily ignored as mental fatigue builds up over the course of an experimental task (Boksem et al., 2005; Faber et al., 2012). Thus, stimulus-driven prioritisation of synchronous audiovisual signals could be more pronounced when mental fatigue has built up.

However, although there were signs of mental fatigue in Experiment 1 and Experiment 4, there was still no evidence for prioritisation of synchronous over asynchronous audiovisual looming signals (Supplementary Materials 4). Thus, even when we consider the effect of time-on-task, we found no evidence in line with our predictions.

Conclusions

The field of multisensory research has mainly focused on studying transient brief audiovisual signals or more complex semantic stimulus information such as audiovisual speech. Using brief audiovisual signals many studies showed attention benefits for multisensory over unisensory signals. However, it remained unclear to what extent these results generalised to multisensory looming signals. In a set of four experiments, we found no evidence for an exogenous nor an endogenous spatial attention benefit for synchronous over asynchronous audiovisual looming signals, nor over unisensory looming signals. In addition, we found no benefits for sustained attention. These experiments further emphasise that attentional benefits for multisensory signals do not always apply to all multisensory signals. Our findings may suggest a lack of early multisensory integration for synchronous looming signals but cannot differentiate this account from an alternative explanation that the signals were not preferentially attended despite being integrated. Our findings further highlight the complexity of the interaction of attention and multisensory integration (Macaluso et al., 2016). For this reason, future multisensory research should focus on computational models to increase conceptual clarity and focus on systematic comparative studies that manipulate multiple factors simultaneously (Innes & Otto, 2019).

Acknowledgements

We would like to thank Lulu Wang for her feedback on this manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work has been funded by the Research Foundation Flanders (FWO) with a grant awarded to HH [1711717N], to CRG [G072517N] and to RvE [G078915N]. Johan Wagemans is funded by the Methusalem program of the Flemish Government [METH/14/02].

Data availability statement

The data, experiment and analysis code used in this study are available on figshare (https://doi.org/10.6084/m9.figshare.9980570.v1).

ORCID

Hanne Huygelier http://orcid.org/0000-0002-5177-1193
Raymond van Ee http://orcid.org/0000-0002-1134-8665
Armien Lanssens http://orcid.org/0000-0002-2791-5556
Johan Wagemans http://orcid.org/0000-0002-7970-1541
Céline R. Gillebert http://orcid.org/0000-0001-6686-7262

References

Abrams, R. A., & Christ, S. E. (2005). The onset of receding motion captures attention: Comment on Franconeri and Simons (2003). *Perception & Psychophysics*, *67*(2), 219–223. https://doi.org/10.3758/BF03206486

Alais, D., van Boxtel, J. J., Parker, A., & van Ee, R. (2010). Attending to auditory signals slows visual alternations in binocular rivalry. *Vision Research*, *50*(10), 929–935. https://doi.org/10.1016/j.visres.2010.03.010

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*(9), 839–843. https://doi.org/10.1016/j.cub.2005.03.046

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637–643. https://doi.org/10.1111/j.1467-9280.2005.01587.x

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, *55*(5), 485–496. https://doi.org/10.3758/BF03205306

Bartolomeo, P., & Chokron, S. (2002). Orienting of attention in left unilateral neglect. *Neuroscience & Biobehavioral Reviews*, *26*(2), 217–234. https://doi.org/10.1016/S0149-7634(01)00065-3

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2005). Effects of mental fatigue on attention: An ERP study. *Cognitive Brain Research*, *25*(1), 107–116. https://doi.org/10.1016/j.cogbrainres.2005.04.011

Boksem, M. A. S., & Tops, M. (2008). Mental fatigue: Costs and benefits. *Brain Research Reviews*, *59*(1), 125–139. https://doi.org/10.1016/j.brainresrev.2008.07.001

Bonnelle, V., Leech, R., Kinnunen, K. M., Ham, T. E., Beckmann, C. F., De Boissezon, X., Greenwood, R. J., & Sharp, D. J. (2011). Default mode network connectivity predicts sustained attention deficits after traumatic brain injury. *Journal of Neuroscience*, *31*(38), 13442–13451. https://doi.org/10.1523/JNEUROSCI.1163-11.2011

Burnham, B. R., & Neely, J. H. (2008). A static color discontinuity can capture spatial attention when the target is an abrupt-onset singleton. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 831–841. https://doi.org/10.1037/0096-1523.34.4.831

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia*, *47*(4), 1045–1052. https://doi.org/10.1016/j.neuropsychologia.2008.11.003

Cecere, R., Romei, V., Bertini, C., & Làdavas, E. (2014). Crossmodal enhancement of visual orientation discrimination by looming sounds requires functional activation of primary visual areas: A case study. *Neuropsychologia*, *56*, 350–358. https://doi.org/10.1016/j.neuropsychologia.2014.02.008

Conrad, V, Bartels, A., Kleiner, M, & Noppeney, U. (2010). Audiovisual interactions in binocular rivalry. *Journal of Vision*, *10*(*10*), 27–27. https://doi.org/10.1167/10.10.27

Conrad, V., Kleiner, M., Bartels, A., O'Brien, J. H., Bülthoff, H. H., & Noppeney, U. (2013). Naturalistic stimulus structure determines the integration of audiovisual looming signals in binocular rivalry. *PLoS One*, *8*(8), e70710. https://doi.org/10.1371/journal.pone.0070710

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron*, *58*(3), 306–324. https://doi.org/10.1016/j.neuron.2008.04.017

Denham, S. L., Farkas, D., van Ee, R., Taranu, M., Kocsis, Z., Wimmer, M., Carmel, D., & Winkler, I. (2018). Similar but separate systems underlie perceptual bistability in vision and audition. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-25587-2

Dent, K., & Humphreys, G. W. (2011). Neuropsychological evidence for a competitive bias against contracting stimuli. *Neurocase*, *17*(2), 112–121. https://doi.org/10.1080/13554794.2010.498381

Einhäuser, W., Methfessel, P., & Bendixen, A. (2017). Newly acquired audio-visual associations bias perception in binocular rivalry. *Vision Research*, *133*, 121–129. https://doi.org/10.1016/j.visres.2017.02.001

Esterman, M., Noonan, S. K., Rosenberg, M., & DeGutis, J. (2013). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cerebral Cortex*, *23*(11), 2712–2723. https://doi.org/10.1093/cercor/bhs261

Faber, L. G., Maurits, N. M., & Lorist, M. M. (2012). Mental fatigue affects visual selective attention. *PLoS One*, *7*(10), e48073. https://doi.org/10.1371/journal.pone.0048073

Farah, M. J., Wong, A. B., Monheit, M. A., & Morrow, L. A. (1989). Parietal lobe mechanisms of spatial attention: Modality-specific or supramodal? *Neuropsychologia*, *27*(4), 461–470. https://doi.org/10.1016/0028-3932(89)90051-1

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(18), 1030–1044. https://doi.org/10.1037//0096-1523.18.4.1030

Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics*, *65*(7), 999–1010. https://doi.org/10.3758/BF03194829

Fujisaki, W., Koene, A., Arnold, D., Johnston, A., & Nishida, S. (2006). Visual search for a target changing in synchrony with an auditory signal. *Proceedings of the Royal Society of London B: Biological Sciences*, *273*(1588), 865–874. https://doi.org/10.1098/rspb.2005.3327

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–285. https://doi.org/10.1016/j.tics.2006.04.008

Gillebert, C. R., Mantini, D., Thijs, V., Sunaert, S., Dupont, P., & Vandenberghe, R. (2011). Lesion evidence for the critical role of the intraparietal sulcus in spatial attention. *Brain*, *134*(6), 1694–1709. https://doi.org/10.1093/brain/awr085

Glatz, C., & Chuang, L. L. (2019). The time course of auditory looming cues in redirecting visuo-spatial attention. *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038/s41598-018-36033-8

Gordon, M. S., & Rosenblum, L. D. (2005). Effects of intrastimulus modality change on audiovisual time-to-arrival judgments. *Perception & Psychophysics*, *67*(4), 580–594. https://doi.org/10.3758/bf03193516

Innes, B. R., & Otto, T. U. (2019). A comparative analysis of response times shows that multisensory benefits and interactions are not equivalent. *Scientific Reports*, *9*(1), 1–10. https://doi.org/10.1038/s41598-019-39924-6

Jaekl, P., Pérez-Bellido, A., & Soto-Faraco, S. (2014). On the "visual" in "audio-visual integration": A hypothesis concerning visual pathways. *Experimental Brain Research*, *232*(6), 1631–1638. https://doi.org/10.1007/s00221-014-3927-8

Johnson, K. A., Kelly, S. P., Bellgrove, M. A., Barry, E., Cox, M., Gill, M., & Robertson, I. H. (2007). Response variability in attention deficit hyperactivity disorder: Evidence for neuropsychological heterogeneity. *Neuropsychologia*, *45*(4), 630–638. https://doi.org/10.1016/j.neuropsychologia.2006.03.034

Johnson, K. A., Robertson, I. H., Kelly, S. P., Silk, T. J., Barry, E., Dáibhis, A., Watchorn, A., Keavey, M., Fitzgerald, M., Gallagher, L., Gill, M., & Bellgrove, M. A. (2007). Dissociation in performance of children with ADHD and high-functioning autism on a task of sustained attention. *Neuropsychologia*, *45*(10), 2234–2245. https://doi.org/10.1016/j.neuropsychologia.2007.02.019

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, *23*, 315–341. https://doi.org/10.1146/annurev.neuro.23.1.315

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One*, *2*(9), e943. https://doi.org/10.1371/journal.pone.0000943

Kucyi, A., Hove, M. J., Esterman, M., Hutchison, R. M., & Valera, E. M. (2017). Dynamic brain network correlates of spontaneous fluctuations in attention. *Cerebral Cortex*, *27*(3), 1831–1840. https://doi.org/10.1093/cercor/bhw029

Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*, *139*(4), 870–900. https://doi.org/10.1037/a0030694

Leo, F., Romei, V., Freeman, E., Ladavas, E., & Driver, J. (2011). Looming sounds enhance orientation sensitivity for visual stimuli on the same side as such sounds. *Experimental Brain Research*, *213*(2–3), 193–201. https://doi.org/10.1007/s00221-011-2742-8

Losier, B. J. W., & Klein, R. M. (2001). A review of the evidence for a disengage deficit following parietal lobe damage. *Neuroscience & Biobehavioral Reviews*, *25*(1), 1–13. https://doi.org/10.1016/S0149-7634(00)00046-4

Macaluso, E., Noppeney, U., Talsma, D., Vercillo, T., Hartcher-O'Brien, J., & Adam, R. (2016). The curious incident of attention in multisensory integration: Bottom-up vs. top-down. *Multisensory Research*, *29*(6–7), 557–583. https://doi.org/10.1163/22134808-00002528

Maier, J. X., Chandrasekaran, C., & Ghazanfar, A. A. (2008). Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Current Biology*, *18*(13), 963–968. https://doi.org/10.1016/j.cub.2008.05.043

Maier, J. X., Neuhoff, J. G., Logothetis, N. K., & Ghazanfar, A. A. (2004). Multisensory integration of looming signals by rhesus monkeys. *Neuron*, *43*(2), 177–181. https://doi.org/10.1016/j.neuron.2004.06.027

Mallick, D. B., Magnotti, J. F., & Beauchamp, M. S. (2015). Variability and stability in the McGurk effect: Contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review*, *22*(5), 1299–1307. https://doi.org/10.3758/s13423-015-0817-4

Marchant, J. L., Ruff, C. C., & Driver, J. (2012). Audiovisual synchrony enhances BOLD responses in a brain network including multisensory STS while also enhancing target-detection performance for both modalities. *Human Brain Mapping*, *33*(5), 1212–1224. https://doi.org/10.1002/hbm.21278

Matusz, P. J., & Eimer, M. (2011). Multisensory enhancement of attentional capture in visual search. *Psychonomic Bulletin & Review*, *18*(5), 904. https://doi.org/10.3758/s13423-011-0131-8

Maxwell, S E, Delaney, H D, & Kelley, K. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). New York: Lawrence Erlbaum Associates Publishers.

McDonald, J. J., Teder-Sälejärvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, *407*(6806), 906–908. https://doi.org/10.1038/35038085

Mcgurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748. https://doi.org/10.1038/264746a0

Mishra, J., Martínez, A., & Hillyard, S. A. (2010). Effect of attention on early cortical processes associated with the sound-induced extra flash illusion. *Journal of Cognitive Neuroscience*, *22*(8), 1714–1729. https://doi.org/10.1162/jocn.2009.21295

Molenberghs, P., Gillebert, C. R., Schoofs, H., Dupont, P., Peeters, R., & Vandenberghe, R. (2009). Lesion neuroanatomy of the sustained attention to response task. *Neuropsychologia*, *47*(13), 2866–2875. https://doi.org/10.1016/j.neuropsychologia.2009.06.012

Moors, P., Huygelier, H., Wagemans, J., de-Wit, L., & van Ee, R. (2014). Suppressed visual looming stimuli are not integrated with auditory looming signals: Evidence from continuous flash suppression. *I-Perception*, *6*(1), 48–62. https://doi.org/10.1068/i0678

Parker, A., & Alais, D. (2007). A bias for looming stimuli to predominate in binocular rivalry. *Vision Research*, *47*(20), 2661–2674. https://doi.org/10.1016/j.visres.2007.06.019

Peirce, J. W. (2007). Psychopy—psychophysics software in python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Piazza, E. A., Denison, R. N., & Silver, M. A. (2018). Recent cross-modal statistical learning influences visual perceptual selection. *Journal of Vision*, *18*(3). https://doi.org/10.1167/18.3.1

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Rengachary, J., d'Avossa, G., Sapir, A., Shulman, G. L., & Corbetta, M. (2009). Is the Posner reaction time test more accurate than clinical tests in detecting left neglect in acute and chronic stroke? *Archives of Physical Medicine and Rehabilitation*, *90*(12), 2081–2088. https://doi.org/10.1016/j.apmr.2009.07.014

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). "Oops!": Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758. https://doi.org/10.1016/S0028-3932(97)00015-8

Rohe, T., & Noppeney, U. (2015). Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biology*, *13*(2), e1002073. https://doi.org/10.1371/journal.pbio.1002073

Rovamo, J., & Raninen, A. (1984). Critical flicker frequency and M-scaling of stimulus size and retinal illuminance. *Vision Research*, *24*(10), 1127–1131. https://doi.org/10.1016/0042-6989(84)90166-4

Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1311–1321. https://doi.org/10.1037/0096-1523.33.6.1311

Santangelo, V., & Spence, C. (2008). Crossmodal attentional capture in an unspeeded simultaneity judgement task. *Visual Cognition*, *16*(2–3), 155–165. https://doi.org/10.1080/13506280701453540

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*(6814), 788. https://doi.org/10.1038/35048669

Shore, D. I., Spence, C., & Klein, R. M. (2001). Visual prior entry. *Psychological Science*, *12*(3), 205–212. https://doi.org/10.1111/1467-9280.00337

Spagna, A., Mackie, M.-A., & Fan, J. (2015). Supramodal executive control of attention. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00065

Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: A review. *Hearing Research*, *258*(1–2), 134–142. https://doi.org/10.1016/j.heares.2009.04.015

Stelmach, L. B., & Herdman, C. M. (1991). Directed attention and perception of temporal order. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(2), 539–550. https://doi.org/10.1037/0096-1523.17.2.539

Stoermer, V. S., Alvarez, G. A., & Cavanagh, P. (2014). Within-hemifield competition in early visual areas limits the ability to track multiple objects with attention. *Journal of Neuroscience*, *34*(35), 11526–11533. https://doi.org/10.1523/JNEUROSCI.0980-14.2014

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. https://doi.org/10.1016/j.tics.2010.06.008

Tarantino, V., Cutini, S., Mogentale, C., & Bisiacchi, P. S. (2013). Time-on-task in children with ADHD: An ex-Gaussian analysis. *Journal of the International Neuropsychological Society*, *19*(7), 820–828. https://doi.org/10.1017/S1355617713000623

ten Oever, S., Romei, V., van Atteveldt, N., Soto-Faraco, S., Murray, M. M., & Matusz, P. J. (2016). The COGs (context, object, and goals) in multisensory processing. *Experimental Brain Research*, *234*(5), 1307–1323. https://doi.org/10.1007/s00221-016-4590-z

Turratto, M., & Pascucci, D. (2016). Short-term and long-term plasticity in the visual-attention system: Evidence from habituation of attentional capture. *Neurobiology of Learning and Memory*, *130*, 159–169. https://doi.org/10.1016/j.nlm.2016.02.010

Tyll, S., Bonath, B., Schoenfeld, M. A., Heinze, H.-J., Ohl, F. W., & Noesselt, T. (2013). Neural basis of multisensory looming signals. *NeuroImage*, *65*, 13–22. https://doi.org/10.1016/j.neuroimage.2012.09.056

Van der Burg, E., Cass, J., Olivers, C. N. L., Theeuwes, J., & Alais, D. (2010). Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS One*, *5*(5), e10664. https://doi.org/10.1371/journal.pone.0010664

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008a). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1053–1065. https://doi.org/10.1037/0096-1523.34.5.1053

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008b). Audiovisual events capture attention: Evidence from temporal order judgments. *Journal of Vision*, *8*(5), 2–2. https://doi.org/10.1167/8.5.2

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2009). Poke and pop: Tactile–visual synchrony increases visual saliency. *Neuroscience Letters*, *450*(1), 60–64. https://doi.org/10.1016/j.neulet.2008.11.002

Van der Burg, E., Talsma, D., Olivers, C. N. L., Hickey, C., & Theeuwes, J. (2011). Early multisensory interactions affect the competition among multiple visual objects. *NeuroImage*, *55*(3), 1208–1218. https://doi.org/10.1016/j.neuroimage.2010.12.068

van Ee, R., van Boxtel, J. J. A., Parker, A. L., & Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *The Journal of Neuroscience*, *29*(37), 11641–11649. https://doi.org/10.1523/JNEUROSCI.0873-09.2009

Walker-Andrews, A. S., & Lennon, E. M. (1985). Auditory-visual perception of changing distance by human infants. *Child Development*, *56*(3), 544–548. https://doi.org/10.2307/1129743

Yu, R. Q., & Zhao, J. (2015). The persistence of the attentional bias to regularities in a changing environment. *Attention, Perception, & Psychophysics*, *77*(7), 2217–2228. https://doi.org/10.3758/s13414-015-0930-5

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science*, *24*(5), 667–677. https://doi.org/10.1177/0956797612460407