BRILL

Multisensory Research 32 (2019) 45–65

MULTISENSORY RESEARCH

brill.com/msr

Cross-Modal Cue Effects in Motion Processing

G. M. Hanada [1,*], **J. Ahveninen** [2,*,**], **F. J. Calabro** [1,3], **A. Yengo-Kahn** [1] **and L. M. Vaina** [1,2,4]

[1] Brain and Vision Research Laboratory, Department of Biomedical Engineering, Boston University, Boston, MA, USA

[2] Harvard Medical School, Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA

[3] Department of Psychiatry and Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA, USA

[4] Harvard Medical School, Department of Neurology, Massachusetts General Hospital and Brigham and Women's Hospital, MA, USA

Received 2 January 2018; accepted 18 September 2018

Abstract

The everyday environment brings to our sensory systems competing inputs from different modalities. The ability to filter these multisensory inputs in order to identify and efficiently utilize useful spatial cues is necessary to detect and process the relevant information. In the present study, we investigate how feature-based attention affects the detection of motion across sensory modalities. We were interested to determine how subjects use intramodal, cross-modal auditory, and combined audiovisual motion cues to attend to specific visual motion signals. The results showed that in most cases, both the visual and the auditory cues enhance feature-based orienting to a transparent visual motion pattern presented among distractor motion patterns. Whereas previous studies have shown cross-modal effects of spatial attention, our results demonstrate a spread of cross-modal feature-based attention cues, which have been matched for the detection threshold of the visual target. These effects were very robust in comparisons of the effects of valid vs. invalid cues, as well as in comparisons between cued and uncued valid trials. The effect of intramodal visual, cross-modal auditory, and bimodal cues also increased as a function of motion-cue salience. Our results suggest that orienting to visual motion patterns among distracters can be facilitated not only by intramodal priors, but also by feature-based cross-modal information from the auditory system.

Keywords

Motion, motion transparency, cross-modal, visual cortex, area MT, audiovisual integration, attention

* These authors made an equal contribution to the work.

** To whom correspondence should be addressed. E-mail: jyrki@nmr.mgh.harvard.edu

DOI:10.1163/22134808-20181313

1. Introduction

In everyday life, we are surrounded by scenes cluttered with many different static and dynamic objects, which cannot be processed simultaneously. Functioning in these types of environments requires attentional mechanisms whose effectiveness is increased by the availability of relevant unimodal and cross-modal sensory cues. Often the unimodal cues are not from the principal modality of the target stimulus, such as when a looming sound guides visual attention to a particular object in a scene that is approaching instead of receding. Much of the previous research orienting of visual attention to motion patterns has, however, concentrated on unimodal studies.

1.1. Cross-Modal Modulation of Motion Perception

Several studies have examined how information from one sensory modality affects the perception of coinciding stimuli in another modality. The effects of cross-modal spatial (Jack and Thurlow, 1973; Kopco *et al*., 2009) or temporal (Shams *et al*., 2000; Shams *et al*., 2002; Vroomen and de Gelder, 2000) cues on stimulus detection have been extensively studied. In addition to the relatively large number of studies regarding spatial or temporal cross-modal influences, several studies have examined cross-modal modulation of the processing of motion direction cues. These studies suggest that cross-modal cues support both auditory and visual motion perception (Cappe *et al*., 2009; Schmiedchen *et al*., 2012; Soto-Faraco *et al*., 2003), and modulate brain activations during motion discrimination tasks (Lewis and Noppeney, 2010; Kayser *et al*., 2017). Our recent studies also demonstrate that a congruent auditory cue can help identify a moving object in an environment where the observer is in self-motion (Roudaia *et al*., 2018), which is highly important for navigation and which underlies several critical activities of daily life, such as collision avoidance. However, overall, the evidence regarding auditory influences on visual motion processing is not entirely consistent. Some studies suggest that while the cross-modal effects are observable in both directions, visual motion cues have larger influences on auditory motion perception than *vice versa* (Bertelson and Radeau, 1981). In the case of apparent motion illusion, one study reported that auditory cues produce no cross-modal effects while visual cues modulate the auditory apparent illusion very clearly (Soto-Faraco *et al*., 2004). Other studies suggest that combined multisensory information direction produces only small additional benefits in the discrimination of linear translational motion (Alais and Burr, 2004). Furthermore, it has been argued that the effects of cross-modal auditory motion cues bias the post-perceptual decision making instead of modulating the sensitivity of visual detection of random dot motion direction, per se (Meyer and Wuerger, 2001).

1.2. Cross-Modal Cueing of Attention to Motion

Previous studies have provided evidence that the effects of attention spread across the different sensory modalities. For example, when a subject attends to a certain visually presented object, responses to auditory stimuli that coincide with that attended object are increased in amplitude (Busse *et al*., 2005; Donohue *et al*., 2011). Whereas the effect of spatial cues is generally stronger from the visual to auditory cortex, auditory-spatial attention cues may be highly useful for visual perception when the potentially significant object is occluded or in the periphery. In such cases, auditory stimuli may provide coarser spatial orientation cues that guide the observers' fine-grained visual attention to the relevant location in a stimulus-driven fashion and subsequently improve the perceptual performance (Driver, 2004; Driver and Spence, 1998; Ward *et al*., 2000).

Given its sensitivity to dynamic patterns, the auditory system could also help subjects orient to visual motion cues. Several studies provide evidence that the expectations of the most likely direction of an object spread not only from visual to auditory domain but also *vice versa* (Beer and Röder, 2004, 2005). These effects were evident in an experimental paradigm where the relative probability of unimodal stimuli was manipulated. However, there is a paucity of studies that have explicitly tested how attended auditory motion cues affect the detection of visual targets amongst several similar competing motion stimuli. Therefore, in the present study, we investigated how subjects use visual, auditory, and audiovisual motion cues to attend to specific visual motion patterns. To accomplish our goal, we used motion-defined transparency as a test case. The transparent motion stimulus was selected based on previous studies, which have addressed its psychophysical (Braddick, 1997; Braddick *et al*., 2002; Calabro and Vaina, 2006; Farell and Li, 2004; Metelli, 1974; Qian *et al*., 1994) and computational (Murakami, 1997; Nowlan and Sejnowski, 1995; Qian *et al*., 1994; Raudies and Neumann, 2011; Raudies *et al*., 2011; Snowden and Verstraten, 1999; Tsai and Victor, 2003; Watanabe and Idesawa, 2003) characteristics, as well as its physiological substrates. The experimental paradigm proposed here provides a solid foundation for investigating how and when unimodal, crossmodal, and bimodal cues affect visual processing at the behavioral and neuronal population levels.

Previous studies show that cross-modal influences may strongly depend on the recent stimulation history and/or training, as evidenced in behavioral (Gondan *et al*., 2004; Otto and Mamassian, 2012), human neuroimaging (Ahveninen *et al*., 2016; Jääskeläinen *et al*., 2007), and monkey neurophysiological studies (Bruns and Roder, 2015). For example, a recent study showed that behavioral indices of non-additive multisensory interactions are not detectable when the stimulus modality changes from trial to trial (Juan *et al*., 2017). On

the other hand, in addition to such bottom-up influences, the potential top-down effects of anticipation in repetitive experimental paradigms (Foxe *et al*., 2014), including in studies of audiovisual attention (Rapela *et al*., 2012), have also been well documented. To control for such effects, we used a paradigm where different cue types were either presented in predictable blocks or in an interleaved design. In addition to the stimulus sequence, we also included control conditions to estimate the effects of cue saliency on motion processing. This allowed us to differentiate effects caused by the actual motion content vs. other cue effects such as alerting. Most importantly, to verify that the subjects benefited from the cue information (in contrast to, e.g., a simple temporal cueing effect), the task trials were interleaved with invalid trials where the cue was incongruent with the target pattern.

Our broader theoretical assumption was that in addition to the intramodal visual cues, auditory motion cues improve the accuracy of feature-based orienting to a visual motion pattern presented among motion distractor patterns. Along these lines, we *specifically hypothesized* that not only for visual, but also for auditory cues, the subjects' performance would be significantly less accurate for incongruent cues than for cues congruent with the target visual motion pattern. Furthermore, we made the specific hypothesis that whereas both visual and auditory motion cues enhance performance in comparison to conditions with no motion cues, bimodal visual-auditory cues lead to better performance, providing stronger behavioral benefits.

2. Material and Methods

2.1. Participants

Twelve observers participated in the study (mean age = 24.75 years, SD = 4; all male). All had normal hearing and normal or corrected-to-normal vision. Three of the observers were authors; the rest of them were naïve as to the purpose of the experiment. All observers gave informed consent according to the Boston University Institution Review board. Prior to enrollment subjects underwent a rigorous practice with the test stimuli for motion detection accuracy (Task 1 with no cues described below). The subjects completed at least three (or up to 10) task blocks of 20 trials to ensure that they could reach an accuracy of at least 60% correct. All subjects were able to achieve the required level of accuracy, most of them in three runs.

2.2. Display and Procedure

Participants were seated at 60 cm viewing distance from the computer monitor in a dark room and were adapted for 5 min to the background luminance of the monitor with head position stabilized with a chin and forehead rest. All

stimuli were generated on a Mac Pro running MATLAB using the BRAVIshell software developed in our laboratory based on the Psychophysical Toolbox (Brainard, 1997; Pelli, 1997), and were presented on a 23′ Apple LCD Cinema Display. All auditory cues were presented with Sennheiser HD201 headphones. We used a Minolta LS-100 light meter for monitor luminance calibration and a Scantek Castle GA-824 Smart Sensor SLM for acoustic calibration.

The visual stimulus consisted of Random Dot Kinematograms (RDK), 50 white dots, 43.9 cd m^{-2} luminance, shown on a gray background (9.9 cd m^{-2}, dot to background contrast was 23%), and moving at 5°/second. The entire screen had the same gray level luminance as the background and the outline of the apertures had the same color and luminance as the RDKs. RDKs were presented in four circular apertures (8° diameter), each displayed within one of four quadrants of the computer screen, and with the center of each aperture being 8° from the fixation mark. Each aperture displayed transparent motion defined by two superimposed RDKs: in one the dots moved horizontally (0° or 180°) and in the other, they moved vertically (90° or 270°). Each RDK was set at 80% coherence such that for any given pair of frames, 80% of the dots were moving in the selected direction (signal dots) while 20% of the dots were repositioned at random locations within the aperture (noise dots). Coherently moving dots were wrapped around the edge of the aperture to maintain a constant density throughout the stimulus duration.

In every trial, the 'target' horizontal direction was randomly selected (e.g., 0° or 180°). One of the four apertures was randomly selected to display the RDK with the horizontally moving dots in the 'target' direction. The direction of horizontal motion in the other three apertures was opposite to the direction of the 'target' (e.g., if the direction target motion direction was 0°, then the horizontal motion in the other three apertures was 180°). In all four apertures there was a simultaneously superimposed second RDK of the same density and luminance as the first, but the motion was vertical (90° or 270°). In each aperture, the specific vertical direction was randomly selected (90° or 270°) and thus did not depend on the horizontally moving RDK in that aperture. Subjects were instructed to selectively attend to the specified motion direction (e.g., horizontal). They were first asked to identify the target aperture (by pressing a predesignated button on the computer keyboard that indicated the quadrant where the target aperture was displayed). Immediately after this, they were asked to report the direction of the vertical motion in the target aperture (by pressing a predesignated button on the computer keyboard).

2.3. Cues

Four types of cues were used to assess the effect of within- and cross-modal attention on facilitating subjects' ability to identify the target aperture. All cue

types provided horizontal motion direction information to the subjects prior to the start of the stimulus in which the horizontal motion direction would match the target horizontal motion direction during the stimulus display. In 20% of the trials ('invalid cue' trials), the horizontal motion direction of the cue was opposite to the direction in the target aperture.

1. No Cue: The aperture presented at the beginning of each trial remained blank and observers performed the task as described above but without having any cue. The task was to identify the target aperture based on the oddball motion direction (e.g., the single aperture (target) with leftward motion, and three other apertures where the horizontal motion was rightward).

2. Visual Cue: A single RDK was displayed inside the center aperture preceding the stimulus presentation. The horizontal motion direction indicated the target's horizontal motion direction. The properties of the RDK were matched to those described above for each aperture, but with only horizontal, and no vertical, motion (i.e., no motion transparency).

3. Auditory Cue: A pure 44.1 kHz tone measured at 83 dB at maximum coherence per ear with 50 dB background noise, travelling either from left to right ear (rightward motion) or right to left ear (leftward motion) by fading the volume from one ear to the other (interaural level differences, ILD) was played through the headphones in the interval preceding the stimulus presentation. During the cue tone, a single blank aperture was presented in the center of the screen for the duration of the cue (300 ms).

4. Combined Cue: Both the visual and the auditory cues described above were presented simultaneously. They always had congruent horizontal motion, so that in the invalid cue trials, both cues were indicating the incorrect direction of motion.

In order to compare the visual and auditory cues, the salience of the cues was adjusted such that the ability to distinguish the presented horizontal motion direction was consistent for each subject. In order to obtain multiple cue levels (threshold and sub-threshold), we used a two-stage procedure for selecting the appropriate difficulty levels: first, an adaptive staircase (Vaina *et al*., 2003) for both visual and auditory cues to determine the approximate threshold, followed by constant stimuli for obtaining a more precise estimation of the psychometric response function. For the visual cues, difficulty was varied by adjusting the motion coherence, e.g., the proportion of dots moving left/right (signal), and the reminder dots (noise) repositioned randomly within the aperture. For the auditory cues, the extent of the ILD was varied, with 100% coherence represented as a tone that started at 100% volume in the

starting ear and 0% volume in the ending ear, and moving to the opposite; for lower coherences, the starting and ending volumes were adjusted such that 0% coherence indicated a sound which did not move at all. In all these conditions, the staircase terminated after 10 reversals, and coherence levels for the last six reversals were averaged to obtain a threshold estimate. This adaptive staircase was repeated three times and averaged across all blocks to determine the subject's threshold detection level. Constant stimulus blocks were then administered with seven coherence levels (at threshold, along with three levels above threshold and three levels below), with 60 trials per level presented in random order. The accuracy vs coherence curve was fit to a two-parameter sigmoid psychometric function. For each subject, the coherence level that corresponded to 76% accuracy ($d' = 1$) was used as a threshold level, and the coherence level that corresponded to 63.8% accuracy ($d' = 0.5$) was used as a subthreshold level for each cue.

2.4. Task

Each trial began with a fixation cross (40 × 40 arcmin; 43.9 cd m^{-2} luminance) displayed in the center of screen. After 300 ms the outline of a single circular aperture was displayed in the center of the screen for 300 ms while the cue was displayed. Notably, in the case of no cue, the aperture was displayed for the same 300 ms. This was followed by a 300 ms stimulus onset asynchrony (SOA) showing only the fixation cross in the center of the screen. It was immediately followed by four transparent motion apertures (as described above) displayed for 1000 ms. The subjects performed a dual task, providing two responses on each trial.

Task 1: subjects were prompted to press a designated key (1–4) on the computer-attached keypad to indicate which of the four quadrants displayed the target aperture (the aperture containing a horizontal RDK with motion opposite to the other three).

Task 2: subjects were asked to report, by pressing the upward or downward arrow on the keypad, the direction of the vertical motion RDK in the target aperture.

Performance was evaluated based on the aperture they selected, even if this was not the correct target aperture, that is, if they identified the wrong target aperture, but correctly identified the motion in that aperture, the second response was considered correct since they were able to identify the motion within the aperture being attended. In Task 1, the reaction times (RT) were recorded relative to the onset of the 1000 ms display of the four apertures containing transparent motion; In Task 2, RTs to each individual trial were

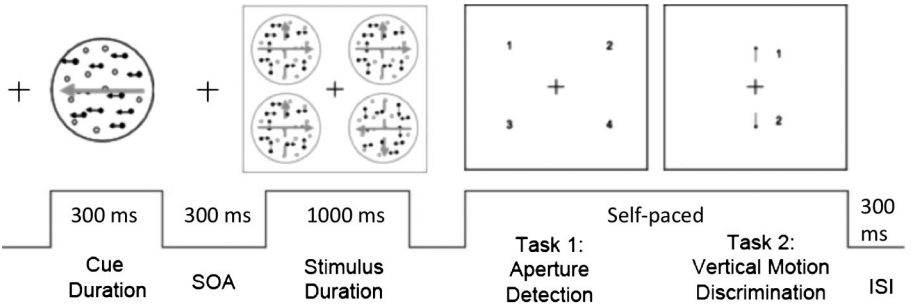

Figure 1. Diagram of a single trial showing the visual cue condition. The cue is presented for 300 ms followed by a 300 ms stimulus onset asynchrony (SOA). Stimulus is then displayed for 1 second. During the response time, in Task 1, observers must detect the target aperture that contains the cued horizontal motion, followed immediately by Task 2, which involves the vertical motion discrimination in the target aperture. There was a 300 ms interstimulus interval (ISI) between all trials.

first baseline corrected by subtracting the RT to Task 1, after which the trial-specific, baseline-corrected RTs were aggregated across all trials in each task condition within each subject. Subjects had unlimited time to enter their responses (Fig. 1). The observers were instructed to respond as accurately as possible. To avoid an impulsive responding strategy, no explicit instruction about the speed of responses was given. There was a 300 ms inter-stimulus interval (ISI) between all trials.

2.5. Block Types

Data was collected in separate block types for each subject: (1) *Uniform*: Twenty trials of each cue type (no cue, visual cue, auditory cue, or combined cues) one at a time. Subjects were told prior to each block which cue type would be provided. Five uniform blocks of each cue type × threshold type combination were collected from each subject, resulting in 100 trials per such a combination in each subject (80 validly cued, 20 invalidly cued). (2) *Interleaved*: Twenty trials of each of the no cue, visual cue, and auditory cue conditions (60 total trials) were presented in a pseudo-randomized interleaved sequence, such that subjects did not know on any given trial which cue type would be presented. Five interleaved blocks of threshold type were collected from each subject, resulting in 100 trials of cue type × threshold type combinations in each subject (80 validly cued, 20 invalidly cued). The test order of each block type and threshold type was randomized for each subject. All subjects completed the task over two days in separate sessions, with half the total number of blocks in each session. In total, the total number of these trials per subject was 1300.

2.6. Statistical Analyses

Given the nature of the task, the main measure of interest was the accuracy of performance, but the reaction times (RT) were also analyzed and reported for completeness. The statistical analyses were conducted using linear mixed effects analyses (LME), by employing the lmer function of the R lme4 module (Bates and Maechler, 2009; Bates *et al*., 2015). The robustness of attention cueing effects was first verified by comparing the effects of valid vs. invalid visual, auditory, and bimodal cues (i.e., this analysis did not consider trials with no cues). The way the different types of cues affected subjects' performance accuracy in the motion processing tasks were analyzed using LME models, which considered the fixed effect of the cue type (visual, auditory, bimodal, or blank aperture), and, additionally, controlled for the fixed effect of cue predictability (uniform vs. interleaved blocks), the interactions between cue and block type, and the random effect of subject identity. The *a priori* comparisons between each individual cue type vs. the no cue condition were derived from these LME models using the 'contr.SAS' function of R. The degrees of freedom for testing the statistical significance were determined using the lmerTest module of R (main effects and interactions were inferred using the 'anova' function, specific *a priori* contrast using the 'summary' function). Finally, RT effects, which are being reported for completeness, were analyzed similarly to the accuracy measures.

3. Results

To verify the effect of intra- vs. cross-modal cues on motion processing, we first compared the subjects' performance after valid vs. invalid visual, auditory, and bimodal motion cues. The way the different types of cues affected subjects' accuracy in the motion processing task were then specified by using LMEs that compared the effects of different cue types on the no cue condition, that is, the blank aperture that provided a temporal cue but no motion-direction information in the 'threshold' condition. In addition to the threshold cues, we conducted a control analysis on the effects of the subthreshold cues.

3.1. The Effect of Cue Validity on Target Detection (Task 1)

To examine the robustness of the cueing effect, and to confirm that the subjects utilized and benefited from the different types of cues, we compared the effects of valid and invalid cues in Task 1. To this end, we utilized an LME analysis that examined the fixed effects of cue validity (valid vs. invalid), cue type (visual cue, auditory cue, or bimodal cue), and block type (uniform vs. interleaved), and which controlled for the random effect of subject identity. The same fixed and random factors were utilized to predict the response accuracy and RT values.

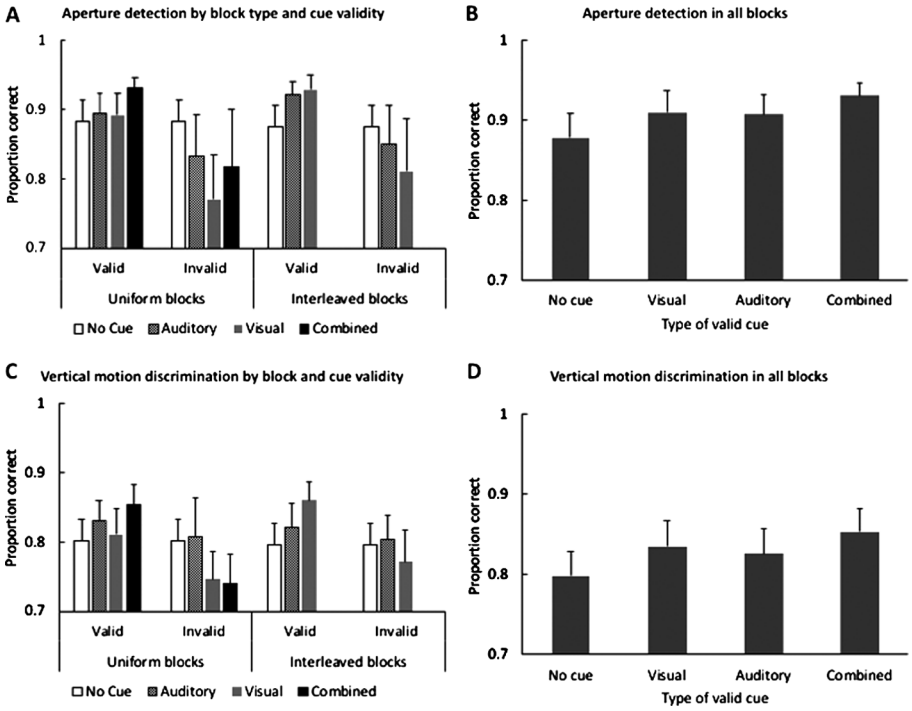

Figure 2. Average motion processing accuracy in the threshold condition across subjects. Error bars reflect the standard error of mean. (A) Aperture detection accuracy in different blocks, for valid vs. invalid cues. (B) Aperture detection (Task 1) performance accuracy in uniform and interleaved blocks in the threshold cue condition. (C) Vertical motion discrimination in different blocks, for valid vs. invalid cues. (D) Vertical motion discrimination (Task 2) performance accuracy after valid cues in uniform and interleaved blocks in the threshold cue condition. Taken together, the data show a significant improvement of visual motion processing accuracy after valid intramodal visual, cross-modal auditory, and bimodal visual-auditory cues. On the other hand, the invalid motion cues deteriorated the performance accuracy, which corroborates that the subjects utilized the cues.

The LME analysis of motion aperture detection *accuracy* showed a highly significant main effect of cue validity [$F(1, 100) = 18.5$, $p < 0.0001$] but no evidence for interactions between the cue validity and cue type or between the cue validity and block type. This verifies that the subjects' performance was significantly more accurate after valid than invalid cues for all cue modality types, irrespective of the predictability of the cue modality. This effect is clearly illustrated in the data shown in Fig. 2A.

The LME analysis of motion aperture detection RTs were consistent with the accuracy analysis: there was a significant main effect of cue validity [$F(1, 100) = 12.2$, $p < 0.001$], but the interaction between cue validity and

Table 1.
Aperture RT data relative to the target onset in different task conditions. For this table, the trial-specific RT values were first aggregated within subjects, after which the group mean and standard deviations (SD) were calculated

Salience	Block type	Validity	Cue type	Mean	SD
Threshold	Uniform	Valid	A	1.13	0.26
			V	1.15	0.30
			VA	1.13	0.33
		Invalid	A	1.21	0.36
			V	1.27	0.40
			VA	1.28	0.46
	Interleaved	Valid	A	1.16	0.30
			V	1.10	0.32
		Invalid	A	1.20	0.34
			V	1.24	0.33
No cue	Uniform			1.14	0.29
	Interleaved			1.19	0.31

cue type and the interaction between cue validity and block type were not statistically significant. In other words, RTs were faster after valid than invalid cues after all cues, irrespectively of the cue predictability (Table 1).

3.2. Effects of Valid Cues on Target Detection (Task 1)

The effect of different cue types on the motion aperture detection was tested using an LME model, which examined the fixed effects of cue type (visual cue, auditory cue, bimodal cue, or blank aperture), block type (uniform vs. interleaved), and the interaction between cue type and block type, and which controlled for the random effect of the subject identity. The same fixed and random factors were utilized to predict the response accuracy and RT values.

The LME analysis of motion aperture detection *accuracy* in Task 1 showed a significant main effect of cue type [$F(3, 66) = 5.6,\ p < 0.01$], whereas the effect of block type and the interaction between cue and block types were non-significant. The *a priori* contrasts computed from the main LME analysis, further, showed that the accuracy of motion aperture detection was significantly higher after visual [$t(66) = 3.1,\ p < 0.01$], auditory [$t(66) = 2.7,\ p < 0.01$], and bimodal [$t(66) = 2.88,\ p < 0.01$] cues than after the blank aperture (Fig. 2B). However, the *a priori* Helmert contrast between the visual and auditory vs. bimodal cues provided no evidence of visual-auditory interactions.

In the LME analysis of *RTs*, the main effects of cue and block type, as well as the interaction between the cue and block types, were all non-significant.

Table 2.
Direction RT data in different task conditions (Task 2). For this table, the trial-specific RTs were first baseline corrected by subtracting respective trial-specific RT to Task 1 and then aggregated within subjects, after which the group means and standard deviations (SD) were calculated

Salience	Block type	Validity	Cue type	Mean	SD
Threshold	Uniform	Valid	A	0.35	0.13
			V	0.36	0.14
			VA	0.37	0.17
		Invalid	A	0.36	0.16
			V	0.35	0.16
			VA	0.35	0.16
	Interleaved	Valid	A	0.36	0.13
			V	0.36	0.14
		Invalid	A	0.36	0.16
			V	0.38	0.14
No cue	Uniform			0.34	0.14
	Interleaved			0.35	0.14

The group averages and standard deviations of RT values in Task 1 are shown in the Table 1.

3.3. The Effect of Cue Validity on Vertical Motion Discrimination (Task 2)

To examine the effect of cue validity, we utilized an LME analysis that examined the fixed effects of cue validity (valid vs. invalid), cue type (visual, auditory, or bimodal), and block type (uniform vs. interleaved), and which controlled for the random effect of subject identity. The same fixed and random factors were utilized to predict the response accuracy and RT values.

The LME analysis of motion discrimination *accuracy* in Task 2 showed a significant main effect of cue validity [$F(1, 100) = 4.6,\ p < 0.001$], but no significant interactions between the cue validity and cue type or between the cue validity and block type. The subjects, thus, performed more accurately after valid than invalid cues after all cue types, irrespective of the cue predictability.

The LME analysis of motion discrimination *RTs* in Task 2 showed no significant effects. The group averages and standard deviations of RT values in Task 2 are shown in the Table 2.

3.4. Effects of Valid Cues on Vertical Motion Discrimination (Task 2)

The effect of motion cues on the vertical motion discrimination was tested using a LME model, which examined the fixed effects of cue type (visual cue, auditory cue, bimodal cue, or blank aperture), block type (uniform vs. interleaved), and the interaction between cue type and block type, and which

controlled for the random effect of the subject identity. The same fixed and random factors were utilized to predict the response accuracy and RT values.

The LME analysis of response accuracy in Task 2 showed a significant main effect of the cue type [$F(3, 66) = 4.6,\ p < 0.01$], but the effect of block type and the interaction between the cue and block types remained non-significant. The *a priori* contrasts computed from the main LME analysis showed that the vertical motion discrimination was significantly more accurate after visual [$t(66) = 3.4,\ p < 0.01$] and bimodal [$t(66) = 2.7,\ p < 0.01$] cues than after the blank aperture. However, the contrast between auditory cue vs. blank aperture was non-significant. The *a priori* Helmert contrast between the visual and auditory vs. bimodal cues suggested no evidence of visual-auditory interactions.

The LME analysis of RTs in Task 2 showed that the main effects of cue or block type, as well as the interaction between the cue and block type, were non-significant.

3.5. Subthreshold Cue Trials

The main purpose of the subthreshold cue conditions was to verify that the potential behavioral effects of intramodal and cross-modal cues are, indeed, related to motion processing instead of nonspecific effects of attention. To this end, we conducted a LME analysis, which examined the fixed effects of cue salience (threshold vs. subthreshold), cue type (visual cue, auditory cue, or bimodal cue), block type (uniform vs. interleaved), and the interactions between cue salience, cue type, block type, and which controlled for the random effect of the subject identity. In the case of Task 1 accuracy, this analysis showed a highly significant main effect of cue salience [$F(1, 99) = 13.4,\ p < 0.001$], that is, the target detection accuracy was significantly better in the threshold vs. subthreshold condition. This provides additional support to our interpretation that the comparisons between threshold cues and the no cue condition (an empty aperture) were not explainable by a simple alerting/temporal cueing effect.

The accuracy values for the subthreshold cues are shown in the Fig. 3, and the RT values in Tables 3 and 4. There was a significant main effect of cue type [$F(3, 66) = 2.8,\ p < 0.05$] in the LME analysis that predicted the Task 1 accuracy as a function of cue type, block type, and their interaction. However, in comparison to the threshold condition that showed significant effects for all cue types vs. the no cue condition, only the effect of visual cues were significant in the *a priori* contrasts in the case of the subthreshold cues [$t(66) = 2.4$, $p < 0.05$]. The main effect of cue type was also significant in the analysis of Task 2 accuracy of subthreshold cues [$F(3, 66) = 3.0,\ p < 0.05$], but the *a priori* contrasts between individual cue types and the no cue condition showed

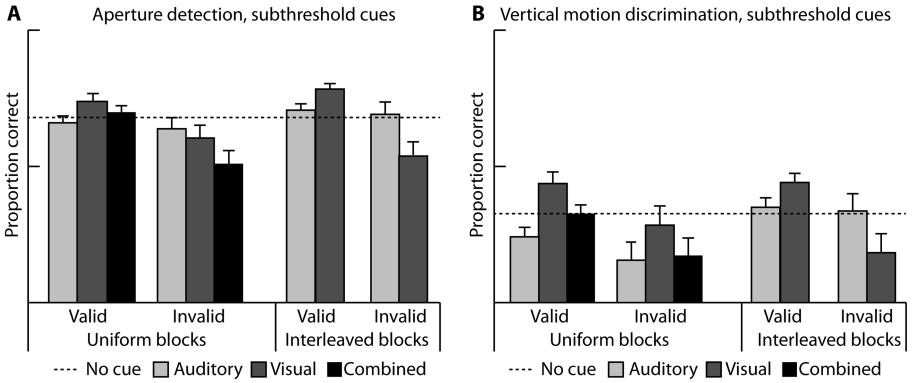

Figure 3. Average performance blocks with subthreshold level ($d' = 0.5$) cues. (A) Aperture detection (Task 1). (B) Vertical motion discrimination (Task 2).

Table 3.
Aperture RT data relative to the target onset in different subthreshold task conditions. For this table, the trial-specific RT values were first aggregated within subjects, after which the group mean and standard deviations (SD) were calculated

Salience	Block type	Validity	Cue type	Mean	SD
Subthreshold	Uniform	Valid	A	1.22	0.38
			V	1.12	0.24
			VA	1.11	0.27
		Invalid	A	1.25	0.36
			V	1.11	0.30
			VA	1.19	0.37
	Interleaved	Valid	A	1.18	0.27
			V	1.11	0.27
		Invalid	A	1.13	0.23
			V	1.13	0.23
No cue	Uniform			1.14	0.29
	Interleaved			1.19	0.31

no significant effects. Finally, all the RT effects in the subthreshold conditions were non-significant.

4. Discussion

In this study, the subjects were presented with visual, auditory, or combined visual-auditory feature cues, which implemented horizontal direction information. The subjects were then asked to identify a target aperture from among four transparent motion apertures, one of which contained a horizontal motion

Table 4.
Direction RT data in different subthreshold cue task conditions. The trial-specific RTs were first baseline corrected by subtracting respective trial-specific RT to Task 1 and then aggregated within subjects, after which the group means and standard deviations (SD) were calculated

Salience	Block type	Validity	Cue type	Mean	SD
Subthreshold	Uniform	Valid	A	0.37	0.16
			V	0.35	0.15
			VA	0.38	0.16
		Invalid	A	0.33	0.10
			V	0.35	0.15
			VA	0.36	0.19
	Interleaved	Valid	A	0.36	0.15
			V	0.35	0.12
		Invalid	A	0.36	0.13
			V	0.35	0.15
No cue	Uniform			0.34	0.14
	Interleaved			0.35	0.14

component moving in the direction of the cue. Importantly, the target aperture could also be identified in the absence of any cue, as the oddball motion pattern among three opposite motion patterns. This suggests that although the cue was not necessary for performing the task, the cues provided an attentional prior for motion direction that facilitated the detection of the target aperture.

By matching the salience of the motion coherence in the visual cues vs. the salience of the inter-aural amplitude gradient in the auditory cues, we were able to determine how a specific cue modality affected the subject's ability to use feature-based attention to motion direction in order to facilitate target detection. The most prominent effect of intramodal and cross-modal cueing was the significant improvement of performance accuracy in the valid cue trials vs. invalid cue trials. This cue validity manipulation was designed based on classic visuospatial attention paradigms, which assume that perceived invalid cues induce additional processing costs due to the need for attention shifting (Posner, 1980; Posner and Petersen, 1990). The same idea was subsequently used in early studies of cross-modal attention, to verify that the subjects indeed benefited from, e.g., auditory cues during orienting of visuospatial attention (Spence and Driver, 1997). In line with these earlier studies, in the present study, invalid cues of all types caused worse performance on aperture discrimination than valid cues. It is also noteworthy that although the performance accuracy was overall the most sensitive measure in our experiments, in the comparisons of valid vs. invalid cues, a significant difference was also observed in RTs. Consistent with recent studies on audiovisual motion

processing (Kayser and Kayser, 2018), the observed performance differences after valid vs. invalid cues provide robust support for our interpretation that the subjects were attending to the cues and were using the cue information when attempting to determine the target aperture.

In more specific comparisons between cued and non-cued trials, we found that valid auditory cues improved the target aperture detection performance almost as much as the equally salient visual cues. This suggests that although visual information on the object's spatial location (Jack and Thurlow, 1973; Kopco et al., 2009) and motion direction (Bertelson and Radeau, 1981; Soto-Faraco *et al*., 2004) overrides simultaneous auditory inputs in perceptual judgments of incongruent multisensory stimuli, in the case of attentional orienting, the effects of motion cues spread also from the auditory to the visual system. This finding is consistent with previous observations in studies of location-specific cross-modal spatial attention, which suggest improved detection of unattended visual targets in auditorily cued locations (Driver and Spence, 1998; Spence *et al*., 1998). However, in contrast to this well-documented spread of spatial attention across modalities, few previous studies have adjusted the cue features themselves to measure the extent of their usefulness. Previously, the spread of feature-based attention has been studied mainly using intramodal visual-only designs (Pessoa *et al*., 2009). Our results extend the previous knowledge of how cross-modal cueing of motion-feature patterns affects target detection.

In the present design, the 'no cue' condition involved an empty visual aperture whose timing matched that of the visual, auditory, and combined cues, to control for temporal cueing influences. However, because the active cues were physically more salient than the empty aperture, the genuine motion-related effects could have been also modulated by temporal cueing influences, as has previously been documented in both intramodal (MacKay and Juola, 2007) and cross-modal (Ten Oever *et al*., 2014) attention studies. This concern is, however, reduced by the comparisons between the threshold and subthreshold cue conditions, which demonstrated that cue-related improvement of aperture detection accuracy is significantly related to the salience of motion information in the cues. This salience effect was, further, most evident in the case of auditory cues, which produced significant accuracy improvements in the threshold but not in the subthreshold cue conditions of Task 1. The fact that the threshold and subthreshold auditory cues differed in the motion salience but not in their average loudness also reduced the likelihood that the threshold cues provided a stronger non-specific alerting effect. Assuming that motion coherence is, in itself, not significantly correlated with non-specific alerting, these notions support our overall interpretation that orienting to visual motion patterns among distracters can be facilitated by feature-based cross-modal information from the auditory system.

In addition to the horizontal motion, all four apertures contained an overlapping vertical motion dot field and subjects were asked to discriminate the direction of vertical motion in the aperture they chose to be the target. In our dual-task design, the subjects were asked to report also the direction of this vertical motion pattern (Task 2). Consistent with the primary task, the significant main effect of our LME model suggested that the cues improved the discrimination accuracy also in this task. However, unlike in Task 1, the effects of cross-modal cues were non-significant. It should be noted, however, that in this secondary task, any cue related effects probably stemmed from more non-specific influences, due to the facilitated detection of the target itself. For example, the subjects could have had more time for motion processing until the target disappeared.

In all analyses, the accuracy effects were statistically more significant than RT measures. The only significant effect in RT was observed in the comparisons of valid vs. invalid trials in Task 1, which showed an effect that was consistent with the accuracy measures: significant improvement of performance for valid vs. invalid cues. The relative insensitivity of RT vs. accuracy measures could reflect a number of factors. For one thing, to avoid an impulsive responding strategy, the subjects were not explicitly instructed to respond as rapidly as possible. The accuracy was the most relevant scientific measure *a priori*. At the same time, given the complexity of the stimulus, the signal-to-noise ratio of RT measures might not have been as good as that of the accuracy measures. It is also possible that some subjects could have preferred to respond only after the offset of the one-second motion pattern, resulting in an 'internal delay' that biased some of the RT measures in these subjects.

A limitation of the present design was that the bimodal cue condition was only included in the uniform block condition. This was done to avoid making the experiment excessively long: As opposed to studies of audiovisual integration, which investigates the effects of exactly or almost simultaneously presented sounds and sights, studies of cross-modal attention typically compare the effects of auditory vs. visual cues that are not necessarily tied to same event, *per se* (e.g., Driver and Spence, 1998). In the present case, the effects of bimodal vs. unimodal cues are slightly more complicated to infer because the improvement could be also explained by the enhancement of cue salience alone, instead of cross-modal or multisensory attention effects. For example, the concurrent auditory stimulus could have increased the perceived salience of the visual motion cue through early direct cross-modal influences from auditory to visual areas (Murray *et al*., 2016; Raij *et al*., 2010; Van Atteveldt *et al*., 2014). However, at the level of orienting that was measured by the accuracy of target detection, we found no evidence of significant multisensory interactions in the 'traditional sense' (i.e., VA $>$ V $+$ A) (Stanford and Stein, 2007), even when tested with the relatively liberal Helmert contrast approach

[VA vs. mean (V,A)], which can be utilized for screening weaker multisensory interactions (Beauchamp, 2005).

In conclusion, whereas previous studies have shown cross-modal effects of spatial attention, our results demonstrate a spread of cross-modal feature-based attention cues, which have been matched for the detection threshold, on visual target detection. These effects were evident in comparisons between cued and uncued conditions, as well as in tasks that compared the effects of valid vs. invalid cues.

Acknowledgements

This work was supported by the National Science Foundation grant 1545668 (LMV), and by the National Institutes of Health grants R01DC016765 (JA), R01DC016915 (JA), and R01MH106174.

References

Ahveninen, J., Huang, S., Ahlfors, S. P., Hämäläinen, M., Rossi, S., Sams, M. and Jääskeläinen, I. P. (2016). Interacting parallel pathways associate sounds with visual identity in auditory cortices, *Neuroimage* **124**, 858–868.

Alais, D. and Burr, D. (2004). No direction-specific bimodal facilitation for audiovisual motion detection, *Brain Res. Cogn Brain Res.* **19**, 185–194.

Bates, D. M. and Maechler, M. (2009). lme4: linear mixed-effects models using S4 classes. *R package version 0.999999-0*.

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015). Fitting linear mixed-effects models using lme4, *J. Stat. Softw.* **67**, 1–48.

Beauchamp, M. S. (2005). Statistical criteria in FMRI studies of multisensory integration, *Neuroinformatics* **3**, 93–113.

Beer, A. L. and Röder, B. (2004). Unimodal and crossmodal effects of endogenous attention to visual and auditory motion, *Cogn. Affect. Behav. Neurosci.* **4**, 230–240.

Beer, A. L. and Röder, B. (2005). Attending to visual or auditory motion affects perception within and across modalities: an event-related potential study, *Eur. J. Neurosci.* **21**, 1116–1130.

Bertelson, P. and Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance, *Atten. Percept. Psychophys.* **29**, 578–584.

Braddick, O. (1997). Local and global representations of velocity: transparency, opponency, and global direction perception, *Perception* **26**, 995–1010.

Braddick, O. J., Wishart, K. A. and Curran, W. (2002). Directional performance in motion transparency, *Vis. Res.* **42**, 1237–1248.

Brainard, D. H. (1997). The psychophysics toolbox, *Spat Vis.* **10**, 433–436.

Bruns, P. and Roder, B. (2015). Sensory recalibration integrates information from the immediate and the cumulative past, *Sci. Rep.* **5**, 12739. DOI:10.1038/srep12739.

Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H. and Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object, *Proc. Natl Acad. Sci. USA* **102**, 18751–18756.

Calabro, F. J. and Vaina, L. M. (2006). Stereo motion transparency processing implements an ecological smoothness constraint, *Perception* **35**, 1219–1232.

Cappe, C., Thut, G., Romei, V. and Murray, M. M. (2009). Selective integration of auditory-visual looming cues by humans, *Neuropsychologia* **47**, 1045–1052.

Donohue, S. E., Roberts, K. C., Grent-'T-Jong, T. and Woldorff, M. G. (2011). The cross-modal spread of attention reveals differential constraints for the temporal and spatial linking of visual and auditory stimulus events, *J. Neurosci.* **31**, 7982–7990.

Driver, J. (2004). Crossmodal spatial attention: evidence from human performance, in: *Cross-Modal Space and Cross-Modal Attention*, Ch. 8, C. Spence and J. Driver (Eds), pp. 179–220. Oxford University Press, Oxford, UK.

Driver, J. and Spence, C. (1998). Cross-modal links in spatial attention, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **353**, 1319–1331.

Farell, B. and Li, S. (2004). Seeing depth coherence and transparency, *J. Vis.* **4**, 209–223.

Foxe, J. J., Murphy, J. W. and De Sanctis, P. (2014). Throwing out the rules: anticipatory alpha-band oscillatory attention mechanisms during task-set reconfigurations, *Eur. J. Neurosci.* **39**, 1960–1972.

Gondan, M., Lange, K., Rosler, F. and Roder, B. (2004). The redundant target effect is affected by modality switch costs, *Psychon. Bull. Rev.* **11**, 307–313.

Jääskeläinen, I. P., Ahveninen, J., Belliveau, J. W., Raij, T. and Sams, M. (2007). Short-term plasticity in auditory cognition, *Trends Neurosci.* **30**, 653–661.

Jack, C. E. and Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the 'ventriloquism' effect, *Percept. Mot. Skills* **37**, 967–979.

Juan, C., Cappe, C., Alric, B., Roby, B., Gilardeau, S., Barone, P. and Girard, P. (2017). The variability of multisensory processes of natural stimuli in human and non-human primates in a detection task, *PLoS One* **12**, e0172480. DOI:10.1371/journal.pone.0172480.

Kayser, S. J. and Kayser, C. (2018). Trial by trial dependencies in multisensory perception and their correlates in dynamic brain activity, *Sci. Rep.* **8**, 3742. DOI:10.1038/s41598-018-22137-8.

Kayser, S. J., Philiastides, M. G. and Kayser, C. (2017). Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations, *Neuroimage* **148**, 31–41.

Kopco, N., Lin, I. F., Shinn-Cunningham, B. G. and Groh, J. M. (2009). Reference frame of the ventriloquism aftereffect, *J. Neurosci.* **29**, 13809–13814.

Lewis, R. and Noppeney, U. (2010). Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas, *J. Neurosci.* **30**, 12329–12339.

Mackay, A. and Juola, J. F. (2007). Are spatial and temporal attention independent?, *Percept. Psychophys.* **69**, 972–979.

Metelli, F. (1974). The perception of transparency, *Sci. Am.* **230**, 90–98.

Meyer, G. F. and Wuerger, S. M. (2001). Cross-modal integration of auditory and visual motion signals, *Neuroreport* **12**, 2557–2560.

Murakami, I. (1997). Motion transparency in superimposed dense random-dot patterns: psychophysics and simulation, *Perception* **26**, 679–692.

Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R. and Matusz, P. J. (2016). The multisensory function of the human primary visual cortex, *Neuropsychologia* **83**, 161–169.

Nowlan, S. J. and Sejnowski, T. J. (1995). A selection model for motion processing in area MT of primates, *J. Neurosci.* **15**, 1195–1214.

Otto, T. U. and Mamassian, P. (2012). Noise and correlations in parallel perceptual decision making, *Curr Biol.* **22**, 1391–1396.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies, *Spat Vis.* **10**, 437–442.

Pessoa, L., Rossi, A., Japee, S., Desimone, R. and Ungerleider, L. G. (2009). Attentional control during the transient updating of cue information, *Brain Res.* **1247**, 149–158.

Posner, M. I. (1980). Orienting of attention, *Q. J. Exp. Psychol.* **32**, 3–25.

Posner, M. I. and Petersen, S. E. (1990). The attention system of the human brain, *Annu. Rev. Neurosci.* **13**, 25–42.

Qian, N., Andersen, R. A. and Adelson, E. H. (1994). Transparent motion perception as detection of unbalanced motion signals. I. Psychophysics, *J. Neurosci.* **14**, 7357–7366.

Raij, T., Ahveninen, J., Lin, F. H., Witzel, T., Jääskeläinen, I. P., Letham, B., Israeli, E., Sahyoun, C., Vasios, C., Stufflebeam, S., Hämäläinen, M. and Belliveau, J. W. (2010). Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices, *Eur. J. Neurosci.* **31**, 1772–1782.

Rapela, J., Gramann, K., Westerfield, M., Townsend, J. and Makeig, S. (2012). Brain oscillations in switching vs. focusing audio-visual attention, *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2012**, 352–355.

Raudies, F. and Neumann, H. (2011). A model of neural mechanisms in monocular transparent motion perception, *J. Physiol. Paris* **104**, 71–83.

Raudies, F., Mingolla, E. and Neumann, H. (2011). A model of motion transparency processing with local center-surround interactions and feedback, *Neural Comput.* **23**, 2868–2914.

Roudaia, E., Calabro, F. J., Vaina, L. M. and Newell, F. N. (2018). Aging impairs audiovisual facilitation of object motion within self-motion, *Multisens. Res.* **31**, 251–272.

Schmiedchen, K., Freigang, C., Nitsche, I. and Rubsamen, R. (2012). Cross-modal interactions and multisensory integration in the perception of audio-visual motion — a free-field study, *Brain Res.* **1466**, 99–111.

Shams, L., Kamitani, Y. and Shimojo, S. (2000). Illusions. What you see is what you hear, *Nature* **408**, 788.

Shams, L., Kamitani, Y. and Shimojo, S. (2002). Visual illusion induced by sound, *Brain Res. Cogn. Brain Res.* **14**, 147–152.

Snowden, R. J. and Verstraten, A. J. (1999). Motion transparency: making models of motion perception transparent, *Trends Cogn. Sci.* **3**, 369–377.

Soto-Faraco, S., Kingstone, A. and Spence, C. (2003). Multisensory contributions to the perception of motion, *Neuropsychologia* **41**, 1847–1862.

Soto-Faraco, S., Spence, C. and Kingstone, A. (2004). Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities, *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 330–345.

Spence, C. and Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting, *Percept. Psychophys.* **59**, 1–22.

Spence, C., Nicholls, M., Gillespie, N. and Driver, J. (1998). Cross-modal links in exogenous covert spatial orienting between touch, audition, and vision, *Atten. Percept. Psychophys.* **60**, 544–557.

Stanford, T. R. and Stein, B. E. (2007). Superadditivity in multisensory integration: putting the computation in context, *Neuroreport* **18**, 787–792.

Ten Oever, S., Schroeder, C. E., Poeppel, D., Van Atteveldt, N. and Zion-Golumbic, E. (2014). Rhythmicity and cross-modal temporal cues facilitate detection, *Neuropsychologia* **63**, 43–50.

Tsai, J. J. and Victor, J. D. (2003). Reading a population code: a multi-scale neural model for representing binocular disparity, *Vis. Res.* **43**, 445–466.

Vaina, L. M., Gryzwacz, N. M., Saiviroonporn, P., Lemay, M., Bienfang, D. C. and Cowey, A. (2003). Can spatial and temporal motion integration compensate for deficits in local motion mechanisms?, *Neuropsychologia* **41**, 1817–1836.

Van Atteveldt, N. M., Peterson, B. S. and Schroeder, C. E. (2014). Contextual control of audio-visual integration in low-level sensory cortices, *Hum. Brain Mapp.* **35**, 2394–2411.

Vroomen, J. and de Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision, *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 1583–1590.

Ward, L. M., Mcdonald, J. J. and Lin, D. (2000). On asymmetries in cross-modal spatial attention orienting, *Percept. Psychophys.* **62**, 1258–1264.

Watanabe, O. and Idesawa, M. (2003). Computational model for neural representation of multiple disparities, *Neural Netw.* **16**, 25–37.