BRILL

Multisensory Research 0 (2021) 1–35

MULTISENSORY RESEARCH

brill.com/msr

Effects of Audiovisual Presentations on Visual Localization Errors: One or Several Multisensory Mechanisms?

Cristina Jordão Nazaré [1,*,**] **and Armando Mónica Oliveira** [2,***]

[1] Instituto Politécnico de Coimbra, ESTESC — Coimbra Health School, Audiologia, 3046-854 Coimbra, Portugal

[2] Faculty of Psychology and Educational Sciences, University of Coimbra, Coimbra, Portugal

Received 20 June 2020; accepted 30 March 2021

Abstract

The present study examines the extent to which temporal and spatial properties of sound modulate visual motion processing in spatial localization tasks. Participants were asked to locate the place at which a moving visual target unexpectedly vanished. Across different tasks, accompanying sounds were factorially varied within subjects as to their onset and offset times and/or positions relative to visual motion. Sound onset had no effect on the localization error. Sound offset was shown to modulate the perceived visual offset location, both for temporal and spatial disparities. This modulation did not conform to attraction toward the timing or location of the sounds but, demonstrably in the case of temporal disparities, to bimodal enhancement instead. Favorable indications to a contextual effect of audiovisual presentations on interspersed visual-only trials were also found. The short sound-leading offset asynchrony had equivalent benefits to audiovisual offset synchrony, suggestive of the involvement of early-level mechanisms, constrained by a temporal window, at these conditions. Yet, we tentatively hypothesize that the whole of the results and how they compare with previous studies requires the contribution of additional mechanisms, including learning detection of auditory-visual associations and cross-sensory spread of endogenous attention.

Keywords

audiovisual integration, spatial localization tasks, auditory cues, crossmodal mechanisms

* To whom correspondence should be addressed. E-mail: cristinanazare@estescoimbra.pt

** ORCID: 0000-0003-4097-1322

*** ORCID: 0000-0002-2308-9904

© Koninklijke Brill NV, Leiden, 2021

DOI:10.1163/22134808-bja10048

UNCORRECTED PROOF

1. Introduction

In recent years, a number of studies on how auditory stimuli can influence visual motion processing have resorted to a representational momentum (RM) paradigm (Chien *et al.*, 2013; Schmiedchen *et al.*, 2012; Teramoto *et al.*, 2010). In the visual RM paradigm, a moving visual target unexpectedly vanishes and the observer is required to indicate its last seen location (for review, see Hubbard, 2005, 2010). The typical result is a forward mislocalization in the direction of motion (Freyd, 1987; Freyd and Finke, 1984). Investigating the modulation of visual RM by sounds may bring together several issues in the realm of audiovisual processing.

First, there is the question of whether auditory information can affect the perceived offset location of a dynamic visual target. Most studies of the impact of sound on visual motion perception have concentrated on motion detection and the processing of motion properties like direction, speed, coherence or phenomenal organization (for a review, see Teramoto *et al.*, 2018), with few examples of crossmodal RM studies (Merz *et al.*, 2020). This may reflect a belief that vision's superiority over audition in spatial localization is even further accentuated with dynamic stimuli (Dick *et al.*, 1987).

Second, there is the question of whether sounds can modulate the perceived location of a moving visual stimulus by means of temporal, spatial, or both kinds of auditory cues (see Teramoto *et al.*, 2012, 2018). Although debated, the temporal coding of a moving object's position is believed to contribute to its spatial location by the visual system (Krekelberg and Lappe, 2001; Whitney, 2002). This opens door to a possible effect of acoustic signals, through the mediation of temporal cues. Indeed, examples have been found wherein the temporal properties of sounds attract those of visual stimuli (temporal ventriloquism: Morein-Zamir *et al.*, 2003), thereby changing their perceived location (Heron *et al.*, 2004; Vroomen and de Gelder, 2004). Regarding spatial auditory cues, it has been shown that the apparent location of auditory tones can attract the perceived location of blurred static blobs (Alais and Burr; 2004; see also Battaglia *et al.*, 2003). However, whether this reverse spatial ventriloquism also extends to moving visual stimuli remains unclear (Kawachi, 2016; Teramoto *et al.*, 2018).

Third, there is the orthogonal question of when added auditory cues can lead to audiovisual (AV) performance enhancement. Sound attraction biases (temporal and reverse spatial ventriloquism) offer only a circumstantial basis for enhancement, depending on the sound pull being in the right direction. For auditory temporal cues, a non-contingent source of AV enhancement might lie in the sharpening of the timing of visual events by sound (but see Huang *et al.*, 2011, and Noesselt *et al.*, 2008). Using a flash-lag setting, Vroomen and de Gelder (2004) showed that while sounds coming before and after the

UNCORRECTED PROOF

flash attracted its temporal occurrence, a synchronized sound sharpened its temporal boundaries and led to a decreased error, consistent with enhanced processing of the flash. Numerous studies similarly documented visual processing enhancement induced by a synchronous sound (e.g., Frassinetti *et al.*, 2002; Li *et al.*, 2015; Lippert *et al.*, 2007; Noesselt *et al.*, 2008, Seitz *et al.*, 2006; Staufenbiel *et al.*, 2011; Stein *et al.*, 1996; Van der Burg *et al.*, 2008; Vroomen and de Gelder, 2000).

Some other studies indicated meanwhile that these beneficial effects could be enabled or disabled by the overall pattern of audiovisual asynchronies. For instance, enhancement of visual contrast detection by a synchronous sound could be cancelled by randomly adding a few AV synchronies (±150 ms, ±400 ms) to the task (Lippert *et al.*, 2007). For related results concerning AV facilitation in visual discrimination and visual detection tasks, respectively, see Li *et al.* (2015) and Targher *et al.* (2017). These results indicate a modulation of the effects of synchronous sounds by the more or less reliable audiovisual temporal contingencies in the setting (Li *et al.*, 2015). Another modulating factor involves perceptual salience of the sound. Vroomen and de Gelder (2000) showed that, in order to enhance visual detection, a synchronous sound had to clearly segregate from a sequence of sounds, producing no beneficial effects if made less abrupt or part of a melody. Comparable findings were reported by Bruns and Getzmann (2008) and Watanabe and Shimojo (2001). These latter authors spoke of context-sensitive saliency-assigning processes, dissociable in principle from perceptual awareness (Watanabe and Shimojo, 2001). Aspects of functional salience were similarly highlighted in the modulatory effects of contextual reliable AV asynchronies (e.g., in temporal AV recalibration, temporal order perception, or AV associative learning), and shown moreover to be modulated by attention (Heron *et al.*, 2010; Hidaka *et al.*, 2015; Ikumi and Soto-Faraco, 2014).

The present study was set out to address the foregoing issues through a set of crossmodal RM tasks (vision as the response-relevant modality; sounds as nontarget stimuli) and a differential comparison with three prior relevant studies, outlined in the following.

Teramoto *et al.* (2010) presented, *via* headphones, non-lateralized sustained sounds that ended before, at, or after visual motion offset (six offset asynchronies in between ±53.2 ms and ±159.6 ms), while participants gazed at a fixation point. The target's perceived location was assessed with a probe judgment method. When sound and motion started synchronously, the forward localization error increased for sounds lasting longer and decreased for sounds lasting less than the visual motion, relative to a silent condition. Chien *et al.* (2013; experiment 1) presented headphone-delivered transient sounds (10 ms) before, at, or after the visual motion ended (six asynchronies in between

UNCORRECTED PROOF

±40 ms and ±140 ms), while participants stared at a fixation cross. Participants gave a mouse-pointing response. Only backward localization errors were observed. Auditory signals at or later than motion offset did not have an effect, but those coming before increased the backward error. Results in both studies were interpreted as an attraction of the perceived timing of visual offset by the timing of sound offset, illustrating a temporally mediated modulation of visual RM.

Chien *et al*. (2013; experiment 2) found that sounds presented to either the left or the right ear produced a stronger backward bias when the ear was on the same spatial side as the origin of motion. After checking with a temporal order judgment (TOJ) task that the perceived temporal relations between sound onset and visual offset were unaffected by side (same *versus* opposite conditions), they concluded to an effect of sound-conveyed spatial information. This rests, however, on evidence from a very different task (TOJ), not furthermore exempt from shortcomings (García-Pérez and Alcalá-Quintana, 2012; Matthews *et al*., 2016). Also, presenting a tone to a single ear (headphone) hardly involves auditory localization cues. General stimulus–response compatibility (Hommel, 2011) or laterality biases toward the stimulated hemispace (Dupierrix *et al*., 2008) could perhaps account for the results.

Schmiedchen *et al*. (2012) generated moving sounds by switching between adjacent loudspeakers, which affords good-quality auditory localization cues (see Meyer *et al*., 2005). Concurrent auditory motion could end spatially behind, at the same point, or ahead of visual motion offset (six spatial offset disparities in between ±5° and ±15° visual angle, v.a.). Participants gazed straight ahead, without a fixation point, and answered by pointing. The perceived offset of short visual motions (500 ms) was shifted toward the endpoints of the sound trajectories, more strongly at the periphery (±60° v.a.) than at the central region (±8° v.a.). Still, significant sound-induced biases occurred in all regions. One problem in interpreting these effects is that spatial and temporal disparities were covaried in the design, meaning it is not possible to disentangle their effects on perceived visual offset location. Importantly, effects were limited to asynchronies within a short critical range (±195 ms). They were absent, for the same spatial disparities, at larger asynchronies associated with a longer visual movement (2000 ms).

One common aspect in all three studies is the lack of beneficial effects of synchronous sounds. Drawing on the reported ability of contextual AV contingencies to enable/disable facilitation by concurrent sounds (Li *et al*., 2015; Lippert *et al*., 2007; Targher *et al*., 2017), we hypothesized that the number (potentially affecting their reliability) and size (potentially affecting their salience) of AV asynchronies might be accountable for that. Accordingly, a distinctive manipulation in our study concerns the number and magnitude of AV temporal disparities. While the comparison studies used six asynchronies,

UNCORRECTED PROOF

always below 195 ms (except for the larger visual motion in Schmiedchen *et al.*, 2012), we used four asynchronies, of (±) 250 ms and 500 ms. These values are considerably above the asynchrony detection thresholds for most types of AV stimuli (see Eg and Behne, 2015; Eg *et al.*, 2015; van Eijk *et al.*, 2008) and similar to those used by Lippert *et al.* (2007), Li *et al.* (2015) and Targher *et al.* (2017). We expected these manipulations to promote relevance of AV disparities to the task and to thereby enable audiovisual benefits by synchronous sounds (and maybe other temporal contingencies). We also expected this to tap into crossmodal mechanisms other than those underpinning sound attraction biases, operating outside the strict time window shown by the comparison studies (see Werner and Noppeney, 2010).

Another distinctive manipulation consisted in combining AV onset and offset disparities (relative to visual motion onset and offset) in a factorial design. This was done in three counterbalanced tasks involving, respectively: (1) temporal AV asynchronies conveyed by non-lateralized continuous sounds (temporal task: T); (2) spatial disparities conveyed by apparently moving sounds (headphone-delivered interaural time differences — ITDs, interaural level differences — ILDs, and average head related transfer function — HRTF, cues), synchronized and matched in duration to the visual motion (spatial task: S); (3) covarying spatial and temporal disparities, obtained through pairing both types of disparities (spatiotemporal task: ST).

The advantage of the onset × offset design is twofold. First, it allows assessing the joint action of the two factors (a point of disagreement between Teramoto *et al.*, 2010 and Chien *et al.*, 2013). Additionally, it plays a methodological role in dissociating the effects of spatial and temporal auditory cues. The use of synchronized stimuli in task S means that concurrent auditory and visual motions of different lengths were achieved using different auditory and visual motion speeds. Since the combination of each auditory spatial offset with different onsets corresponds to different auditory motion speeds, a null effect of onset will exclude an influence of sound speed. If a significant effect of onset is found, the factorial structure of the design should still allow assessing the effect of auditory spatial offset *via* its possible interactions with sound speed.

Participants in our study answered through pointing with a mouse (the typical response with continuous motion targets: Hubbard, 2005) and were allowed to freely move their eyes. For different reasons, all comparison studies asked subjects to keep their eyes still. In Schmiedchen *et al.* (2012) this was aimed at contrasting results from spatial offset disparities in central vision and at the periphery, where spatial reliability of audition relative to vision increases (Hidaka *et al.*, 2011; Perrott *et al.*, 1993). While stronger at the periphery, modulation by sound was also present in central vision. However, this could be due to the covarying AV temporal disparities in the study (effects of sound

UNCORRECTED PROOF

temporal cues in the other comparison studies were established with motion offsets less than ±3° v.a. around the midline). By leaving eye movements unconstrained, along with a more strict dissociation of temporal and spatial disparities, we sought to clarify whether auditory spatial cues could also impact visual RM under foveal vision. Additionally, RM for smooth visual motion has been shown to decrease, and even be eliminated, when oculomotor tracking is prevented (De Sá Teixeira *et al.*, 2013; Kerzel, 2000; Stork and Müsseler, 2004), which seems like a less than optimal property for the purpose of investigating its modulation by sound.

2. Method

2.1. Participants

Thirty-one undergraduate students of the University of Coimbra and the Coimbra Health School (25 female, six male), with ages ranging from 18 to 22 years ($M = 19.35$, SD = 1.603), volunteered to the study. Twenty-one received course credits for their participation. They were all naïve of the purpose of the experiments. None of them reported any type of hearing loss and all had normal or corrected-to-normal vision. Their informed consent was obtained and all aspects of procedure complied with the Helsinki Declaration.

2.2. Stimuli

Stimuli consisted of.avi animations (at 60 frames per second, fps) including both visual-only (silent: V) and audiovisual (coupled with sound: AV) presentations of a light gray disk of 1 cm diameter (≈35 pixels at the used screen resolution; ≈0.7° at a viewing distance of 84.5 cm), moving horizontally from left to right at a constant velocity (8 cm/s ≈ 5.4° v.a./s) and suddenly vanishing at a variable point on the screen. This velocity is within the lower range of velocities used in studies of visual RM (see De Sá Teixeira and Oliveira, 2014; De Sá Teixeira *et al.*, 2013; Hubbard, 1990; Hubbard and Bharucha, 1988). Regarding auditory motion, lower velocities (as long as exceeding 1° v.a./s: Perrott and Saberi, 1990) can in general be expected to ease the localization of dynamic sound targets (Carlile and Leung, 2016; Chandler and Grantham, 1992; Perrott and Musicant, 1977), a desirable property in the current study. Visual-only stimuli were the same in all tasks. The audio streams of the AV stimuli, on the other hand, comprised three distinct sets, one for each task (T, S and ST). Sounds were generated using Audacity v2.0.2 (https://audacityteam.org/) and consisted of band-passed filtered white noise (2 kHz bandwidth centered at 2 kHz, sampling frequency of 44.1 kHz, 32-bit floating depth, 15 ms linear fade-in and fade-out). This bandwidth is commensurate with the bandpass frequency range of 1–3 kHz used in Getzmann's (2005) study of the auditory Fröhlich effect, and in Getzmann *et al.*'s (2004)

UNCORRECTED PROOF

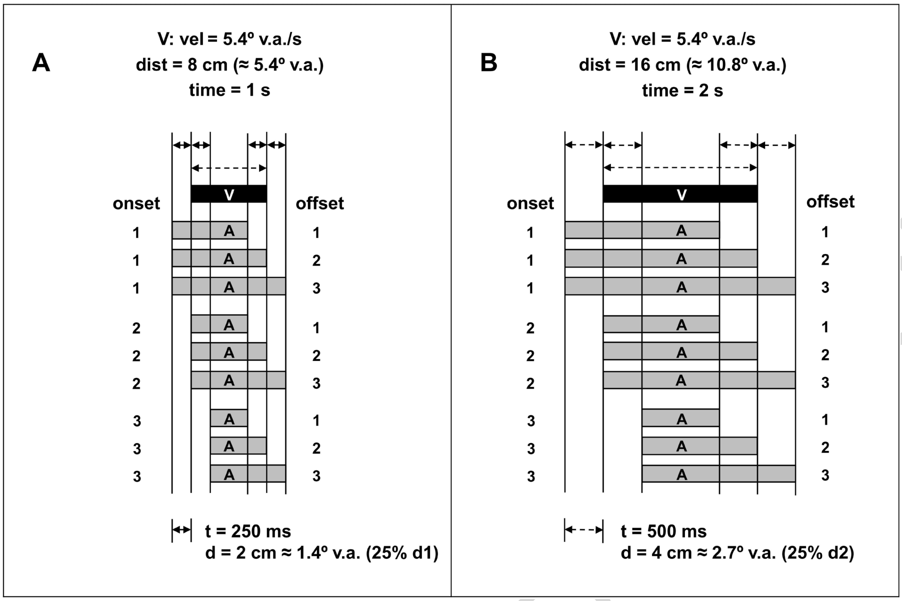

Figure 1. Schematic diagram of the experimental design. The two black rectangles in panels A and B represent two visual motions of different length/duration. The two sets of gray rectangles placed below them stand for durations of or distances travelled by auditory stimuli. In each panel, misalignments between the endpoints of the gray and the black rectangles represent onset and offset audiovisual disparities (which may be temporal, spatial, or both). Each panel illustrates a 3 (AV onset disparities) × 3 (AV offset disparities) design associated with a given visual motion.

and Getzmann and Lewald's (2007, 2009) studies of auditory RM. Unlike the localization of tonal and narrow-band stimuli, which becomes less accurate at middle frequencies centered on 2 kHz (Middlebrooks and Green, 1991), localization accuracy of broadband filtered noise spanning two octaves or more is not dependent on center frequency, at least for normal hearing participants (Risoud *et al.*, 2018; Yost and Zhong, 2014; Yost *et al.*, 2013). Sound durations varied from 0.5 s to 3 s, in accordance with the experimental design (see Fig. 1). These sound durations were derived as per the next subsection from the selected durations of visual motion (1 s and 2 s), which fall within a range covered by previous studies (e.g., Feinkohl *et al.*, 2014; Schmiedchen *et al.*, 2012; Teramoto *et al.*, 2010). Sound spatialization, required in tasks S and ST, was performed with slab3d v6.6.1 (http://slab3d.sonisphere.com/) using ITDs, ILDs and an average HRTF (Miller and Wenzel, 2002) for presentation over headphones. Spatial sounds in task S travelled at constant velocities of ≈2.7°, 4.1°, 5.4°, 6.8° and 8.2° v.a./s, in order to cover different distances within the same durations as the visual motion. All sounds were saved as stereo.wav files and merged with video streams using VirtualDub v1.9.11 (http://virtualdub.

UNCORRECTED PROOF

sourceforge.net/). The duration of each video extended up to 500 ms after the vanishing of the moving gray disk (i.e., up to 30 frames after the last frame in which the disk was presented).

2.3. Experimental Design

The diagrams in Fig. 1A, B illustrate the logic of the experimental design, valid for all tasks. Visual motion is represented by the black rectangles. The black rectangle in (A) represents a distance of 8 cm $\approx$ 5.4° v.a. covered by the moving visual target or, equivalently, a travel duration of 1 s at a constant velocity of 5.4° v.a./s (d1). The black rectangle (B) represents a distance covered of 16 cm $\approx$ 10.8° v.a. or, equivalently, a travel duration of 2 s at a constant velocity of 5.4° v.a./s (d2). Similar distances/durations have been used in studies of visual RM (De Sá Teixeira and Oliveira, 2014; De Sá Teixeira *et al.*, 2013; Hubbard and Lange, 2010; Hubbard and Ruppel, 1999). The gray rectangles in both panels stand for sound durations (non-lateralized sounds of task T) or sound travelled distances (spatialized sounds in tasks S), showing the relative disparities of auditory onset/offset with respect to visual motion onset/offset. The represented disparities can be temporal (task T), spatial (task S), or spatial and temporal (task ST). In all cases, they correspond to a constant fraction ($\pm$25%) of the duration/travel distance of the visual motion. Hence, AV onset/offset asynchronies were of −250, 0, +250 ms for visual motions of 1 s duration (d1), and of −500, 0, +500 ms for visual motions of 2 s duration (d2). Likewise, AV spatial disparities at the onset/offset were of −2, 0, +2 cm ($\approx$1.4° v.a.) for visual motions of 8 cm (d1) and of −4, 0, +4 cm ($\approx$2.7° v.a.) for motions of 16 cm (d2). In task ST, the two kinds of disparities were presented in a covarying manner (i.e., temporal and spatial AV disparities of the same signal and proportional size in relation to the travel distance/duration of the visual motion were paired: −250 ms and −1.4° v.a.; −500 ms and −2.7° v.a.; +250 ms and +1.4° v.a.; +500 ms and +2.7° v.a.).

Not shown in Fig. 1, each distance covered by the visual target was associated with two distinct target vanishing points, 3 cm ($\approx$2° v.a.) to the left and 5 cm ($\approx$3.4° v.a.) to the right of the horizontal midline of the screen. This ensures that both targets moving toward and away from the center are included (see Hubbard and Ruppel, 1999; Munger *et al.*, 2006), while avoiding an exact symmetry that could act as a positional information to the participants. These values are close to the 2° to 3° v.a. left or right from the center used for vanishing positions in Teramoto *et al.* (2010), and to the $\pm$0.3° v.a. jitter around the midpoint used by Chien *et al.* (2013). Finally, considering the size of spatial AV disparities in the design, they entail that spatial auditory offsets spanned a range of positions from $\approx$−4.8° v.a. to the left of the center to $\approx$6.1° v.a. to the right, never approaching the right edge of the screen ($\approx$13.9° v.a. from the center). The design in each task thus corresponded to a 3 (auditory onset:

UNCORRECTED PROOF

hereafter, onset) × 3 (auditory offset: hereafter, offset) × 2 (distance/duration: hereafter, distance) × 2 (vanishing point: hereafter, point) within-subjects factorial design. Each condition defined by the onset × offset factorial crossing thus occurred four times. In addition to the 36 AV presentations corresponding to the main design, each task included two extra sets of trials: (1) four presentations of visual motion alone (two travel distances with two vanishing points each), allowing for comparisons between unimodal and bimodal conditions: these will be denoted as V_T, V_S and V_ST when distinguishing between their AV contexts is needed; (2) 24 V and AV presentations exhibiting a variety of vanishing points and travel distances, all distinct from those in the main design trials, to avert participants from predicting the motion endpoints. These trials included 3 V and 21 AV displays (seven of which without AV disparities), with eight distinct vanishing points and three distinct travel distances randomly assigned to the 24 trials. Responses to these control trials were not recorded. Participants carried out the three tasks in different sessions, separated by a few days ($M = 4.35$, SD $= 1.99$, in between the first and second sessions; $M = 4.45$, SD $= 1.65$, in between the second and third sessions). Tasks were counterbalanced across participants using all six possible orderings, as an additional control for order effects. By taking tasks T, S, and ST as a further within-subjects factor, the whole study can finally be described as a repeated-measures five-factor full factorial design: 3 (task) × 3 (onset) × 3 (offset) × 2 (distance) × 2 (point).

Participants were further invited to do an ancillary task that took place about one month after the last session (in days: $M = 34.38$, SD $= 2.60$), consisting in locating the vanishing point of a moving sound (auditory-only conditions: A). Stimuli were a subset of the audio streams used in task S, limited to conditions where auditory motion had the same spatial onset as visual motion (onset 2 in panels A and B of Fig. 1) and ended at each of the three assigned spatial auditory offsets (offsets 1, 2 and 3 in panels A and B of Fig. 1). This task was devised to test the localizability of the auditory offset points. All 31 participants in the main experiments attended the ancillary task.

2.4. Apparatus and Procedure

Animations were randomly presented on a flat LCD monitor (Samsung, SyncMaster 940BW; screen resolution 1440 × 900 pixels, corresponding to about 27.8° × 13.9° v.a.; refreshing rate: 60 Hz; Samsung). Auditory stimuli were delivered by headphones (Sennheiser HD 202; range of frequency response: 18 Hz–18 kHz; Sennheiser Electronic GmbH & Co. KG, Wedemark, Germany) at approximately 60 dB (A), measured at the output with a sound level meter (Brüel & Kjær 2260; Brüel & Kjær, Nærum, Denmark). All aspects of stimuli presentation and response collection were controlled by SuperLab v4.0.7b (https://cedrus.com/superlab/). Experiments took place in darkness, in

UNCORRECTED PROOF

a soundproofed room. Participants sat comfortably at a distance of 84.5 cm from the screen, heads resting on a chin–forehead support that aligned their eyes with the vertical and horizontal midlines of the screen. A single-channel recording of the word ‘attention’ was presented diotically 1000 ms before each stimulus, without participants being asked to fixate any specific point or region of the screen (to avoid landmark effects: Hubbard and Ruppel, 1999). No restrictions on eye movements during the task were imposed. Participants were instructed to place a ‘plus sign’ cursor, controlled by a wireless mouse, over where they perceived the center of the gray disk to be when they last saw it and click on a mouse button. Compared to probe judgment, mouse-pointing affords a more direct measure of localization (with pixel-level precision) and requires much fewer trials (Hubbard, 2005). It also produces more substantial displacements (Kerzel and Gegenfurtner, 2003), which here is a convenient feature. Instructions made no mention of the auditory stimuli other than that participants should act the same irrespective of whether the visual target was presented alone or accompanied by a sound. The cursor became visible at the center of the screen 500 ms after each presentation, 4 cm below the vertical midline. Following an answer, the message ‘click to continue’ was displayed at the center of the screen until the participant triggered the next trial with a mouse button press. Localization responses were recorded as XY screen coordinates in pixels. Participants always started by completing a practice run of 10 trials randomly selected among the experimental trials.

2.5. *Data Analysis*

Graphical and statistical analyses, consisting mainly of repeated-measures analyses of variance (ANOVAs), were based on horizontal localization errors, computed as the difference between the x-coordinate of the point indicated by the participant and the x-coordinate of the true final location of the target on the screen. Errors are presented hereafter as angular displacements (in degrees of v.a.). Positive errors signal a forward shift of the target’s perceived location, negative errors a backward shift. Following Hubbard (1995a, b), we refer to horizontal errors as M displacements (along the axis of motion). A p value of 0.05 was taken as the significance threshold. Greenhouse–Geisser epsilon correction was used whenever sphericity was violated (assessed with Mauchly’s test). All follow-up pairwise comparisons were Bonferroni-corrected for Type-1 error inflation. Comparisons of variances were performed with the Pitman–Morgan test for correlated samples (Morgan, 1939; Pitman, 1939; Snedecor and Cochran, 1989).

UNCORRECTED PROOF

3. Results

3.1. Overall Five-Way Repeated-Measures ANOVA

Table 1 presents the results of the overall 3 (task) × 3 (onset) × 3 (offset) × 2 (distance) × 2 (point) repeated-measures ANOVA. All factors except onset

Table 1.
Five-way repeated-measures ANOVA: 3 (task) × 3 (onset) × 3 (offset) × 2 (distance) × 2 (point). Starting with four-way interactions, Onset, Offset and Distance are abridged to On, Off and Dist, respectively. Reported p values were Greenhouse–Geisser-corrected when the sphericity assumption was violated. For simplicity, the table presents uncorrected df values for those cases, along with the Greenhouse–Geisser epsilon (ε) used for correction (rightmost column in the table). Significant values ($p \leqslant 0.05$) are marked with an asterisk

Factors	df_{effect}	df_{error}	F	p	${\eta_p}^2$	ε
Task	2	60	4.11	0.032*	0.12	0.768
Onset	2	60	0.06	0.942	<0.01	–
Offset	2	60	10.38	<0.001*	0.26	–
Distance	1	30	20.00	<0.001*	0.40	–
Point	1	30	12.13	0.002*	0.29	–
Task × Onset	4	120	0.70	0.560	0.02	0.792
Task × Offset	4	120	0.60	0.662	0.02	–
Task × Distance	2	60	3.96	0.024*	0.12	–
Task × Point	2	60	0.41	0.538	0.01	0.532
Onset × Offset	4	120	0.26	0.904	<0.01	–
Onset × Distance	2	60	1.34	0.271	0.04	–
Onset × Point	2	60	1.73	0.192	0.06	0.802
Offset × Distance	2	60	11.59	<0.001*	0.28	–
Offset × Point	2	60	0.95	0.392	0.03	–
Distance × Point	1	30	10.02	0.004*	0.25	–
Task × Onset × Offset	8	240	1.08	0.375	0.04	–
Task × Onset × Distance	4	120	0.33	0.799	0.01	0.739
Task × Onset × Point	4	120	0.83	0.506	0.03	–
Task × Offset × Distance	4	120	2.60	0.046*	0.08	–
Task × Offset × Point	4	120	2.97	0.022*	0.09	–
Task × Distance × Point	2	60	1.37	0.262	0.04	–
Onset × Offset × Distance	4	120	0.817	0.517	0.03	–
Onset × Offset × Point	4	120	1.09	0.366	0.04	–
Onset × Distance × Point	2	60	0.683	0.477	0.02	0.788
Offset × Distance × Point	2	60	1.325	0.273	0.04	–
Task × On × Off × Distance	8	240	1.182	0.321	0.038	0.623
Task × On × Off × Point	8	240	1.265	0.263	0.04	–
Task × Off × Dist × Point	4	120	1.169	0.321	0.038	–
Task × On × Dist × Point	4	120	0.231	0.921	0.008	–
On × Off × Dist × Point	4	120	1.568	0.187	0.05	–
Task × On × Off × Dist × Point	8	240	0.643	0.741	0.021	–

UNCORRECTED PROOF

($p = 0.942$) yielded significant main effects. There were three two-way significant interactions, all involving distance: task × distance ($p < 0.05$), offset × distance ($p < 0.001$), distance × point ($p < 0.01$). The only remaining significant results were two three-way interactions: task × offset × distance ($p < 0.05$) and task × offset × point ($p < 0.05$).

The absence of effects of onset (either main or interaction) indicates that AV disparities were only operative at the offset. The significant main effect of task ($p < 0.05$) indicates that the type of auditory information (temporal, spatial or covarying spatial and temporal disparities) influenced the magnitude of M displacement. The statistically null task × offset interaction ($p = 0.662$) indicates moreover that this did not alter the way offset disparities operated across tasks.

Distance and vanishing point are less theoretically relevant variables (given the problems addressed), the main interest of which lies in how they potentially modulate the remaining factors. The fact of both having significant main effects (respectively, $p < 0.001$ and $p = 0.01$) supports that they are inherently distinct variables, against indications that distance effects might suffer from a methodological confound with vanishing point (De Sá Teixeira and Oliveira, 2011). Since the magnitude of AV disparities varied as a fixed percentage (25%) of the distance travelled by the visual target (for spatial disparities) or, equivalently, of its time of travel (for temporal disparities), the significant task × distance and offset × distance interactions may signal a modulatory effect of the size of offset disparities upon both task and offset.

These results are further analyzed and broken down in the following sections with the help of graphical analyses and, as appropriate, more focused comparisons addressing subsets of data in the design.

3.2. *Effects of Audiovisual Offset Disparities*

Figure 2 presents the factorial plot of auditory onset × auditory offset averaged across all other factors. Nearly flat lines indicate no effect of onset (on the abscissa), the vertical separation of lines the effect of offset (curve parameter), and parallelism in the graph that there is no interaction between the two factors. All this was statistically buttressed in the overall ANOVA (Table 1).

The null interaction task × onset × offset in Table 1 ($p = 0.375$) indicates that this factorial pattern remained stable across tasks, not being an aggregation artifact. Effects of offset are vertically ordered, by increasing M displacement, as off_2 (no AV offset disparities), off_1 (sound offset spatially and/or temporally prior to visual offset) and off_3 (sound offset temporally later and/or spatially ahead of visual offset). Follow-up pairwise comparisons disclosed a significant difference between off_2 and off_3 ($p < 0.001$), but not between off_1 and off_2 ($p = 0.09$) or off_1 and off_3 ($p = 0.129$). The null task × offset interaction in Table 1 ($p = 0.662$) indicates furthermore that

UNCORRECTED PROOF

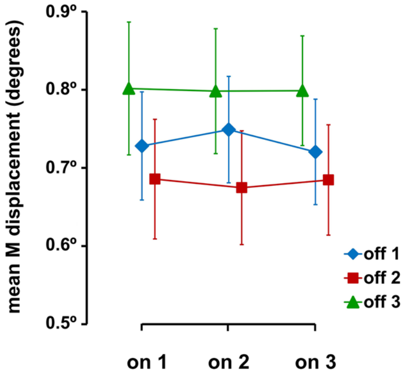

Figure 2. Factorial plot of auditory onset × auditory offset pooled across all the remaining factors. Mean M displacement is given on the ordinate. Error bars stand for standard error of the mean (SEM). Since data are aggregated over tasks (T, S and ST), onset and offset refer here indistinctly to temporal, spatial, and pairs of covarying spatial and temporal disparities.

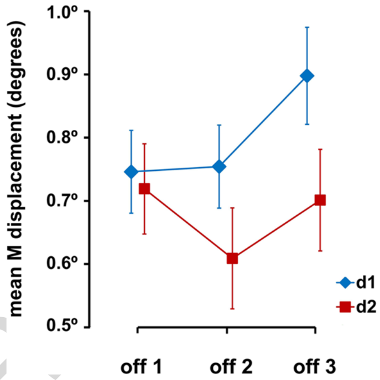

Figure 3. Factorial plot of auditory offset × distance (d1, d2) averaged across all the remaining factors. Mean M displacement values are on the ordinate. Error bars stand for SEM.

this ordering is not due to averaging and remains the same across tasks. Subsequent pairwise comparisons between levels of offset carried out in each task revealed the same exact statistical pattern as in the overall analysis, with significant differences between off_2 and off_3 (T: $p < 0.01$; S: $p = 0.05$, ST: $p = 0.01$), and no significant differences between off_1 and off_2 or off_1 and off_3 (lowest observed $p = 0.160$, for off_1 *versus* off_2 in task T; highest observed $p = 0.933$, for off_1 *versus* off_2 in task ST).

3.3. *Effects of Audiovisual Disparities Magnitude*

Figure 3 illustrates the significant offset × distance interaction signaled in Table 1 ($p < 0.001$). This interaction involves the relative location of off_1

UNCORRECTED PROOF

regarding off_2 and off_3 in the shorter (d1) and in the larger (d2) travelled distances. AV disparities are proportionally smaller (half the size) at d1 (±250 ms/±1.36°) than at d2 (±500/±2.71°). Thus, with small AV disparities off_1 produces the same M displacement [$M = 0.72°$; ±0.06° standard error of the mean (SEM)] as off_2 ($M = 0.73°$, ±0.06° SEM), and both depart from off_3 ($M = 0.87°$, ±0.07° SEM). With large AV disparities, in turn, off_1 produces similar M displacement ($M = 0.69°$, ±0.07° SEM) to off_3 ($M = 0.68°$, ±0.08° SEM) and both depart from off_2 ($M = 0.59°$, ±0.08° SEM). This was statistically supported by follow-up pairwise comparison showing, in d1: a null result for off_1 *versus* off_2 ($p = 1.0$); significant results for off_3 *versus* off_1 ($p < 0.001$) and *versus* off_2 ($p < 0.01$); in d2: a null result for off_1 *versus* off_3 ($p = 1.0$); significant results for off_2 *versus* off_1 ($p = 0.001$) and *versus* off_3 ($p < 0.01$).

The two levels of distance were also related to different sound durations, larger in d2. Differences in total sound duration, rather than in size of disparities, might thus underlie the effect of distance on offset. This is however excluded by the null effect of onset, as onset variation entailed a variation in duration. This interaction can thus be interpreted as a modulation of the impact of AV offset disparities by their size.

A further significant result in Table 1, concerning the three-way interaction task × offset × distance ($p < 0.05$), signals that this modulation actually depends on the task. Figure 4 presents the factorial plot offset × distance separately for each task, allowing seeing that the interaction remains in tasks T and ST, involving temporal disparities, but vanishes in task S, involving only spatial disparities.

These indications from visual inspection where statistically supported by repeated-measures ANOVAs conducted separately for each task. Significant

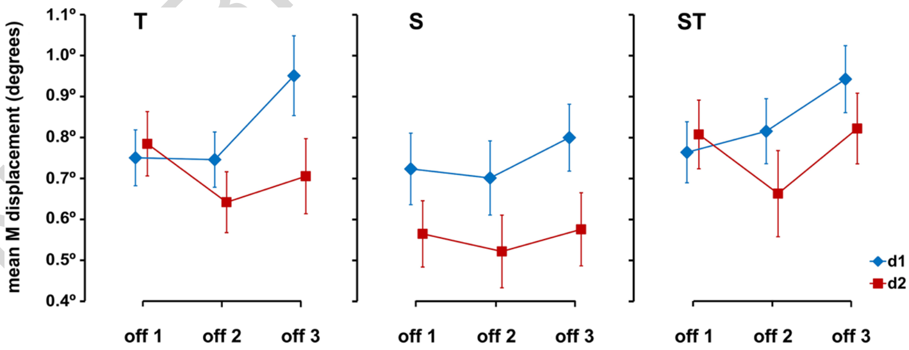

Figure 4. Factorial plots of auditory offset × distance (d1, d2) in each task. T: Temporal task (temporal disparities only). S: spatial task (spatial disparities only). ST: spatiotemporal task (covarying temporal and spatial disparities). Mean M displacement is shown on the ordinate. Error bars denote SEM.

UNCORRECTED PROOF

results for the offset × distance interaction emerged in tasks T, $F_{2,60} = 8.280$, $p = 0.001$, $\eta_p{}^2 = 0.216$, and ST, $F_{2,60} = 6.057$, $p < 0.01$, $\eta_p{}^2 = 0.168$, but not in task S, $F_{2,60} = 0.809$, $p = 0.460$, $\eta_p{}^2 = 0.026$. Subsequent pairwise comparisons carried out in tasks T and ST revealed the same statistical pattern as in the aggregated analysis: in d1, null results for off_1 *versus* off_2 (T: $p = 1.0$; ST: $p = 0.912$) and significant results for off_1 *versus* off_3 (T: $p < 0.01$; ST: < 0.05); conversely, in d2, null results for off_1 *versus* off_3 (T: $p = 0.515$; ST: $p = 1.0$) and significant results for off_1 *versus* off_2 (T: $p < 0.01$; ST: $p < 0.05$).

Taken together, these results indicate that it is the size of temporal AV offset disparities (asynchronies), not of spatial AV disparities (spatial misalignments) that drives the offset × distance interaction, absent in task S. They suggest further an asymmetric temporal window within which the small −250 ms AV offset asynchronies are treated as equivalent to synchronous AV offsets.

3.4. *Effects of Task: Temporal, Spatial and Spatiotemporal Audiovisual Disparities*

Figure 5 presents the onset × offset plots corresponding to each task (from left to right: T, S and ST). All plots show, with only more irregularity, the same parallelism trend noted in the overall ANOVA. Separate repeated-measures ANOVAs performed for each task disclosed in all cases a null effect of onset, T: $F_{1.644,49.307} = 0.890$, $p = 0.416$ ($\varepsilon = 0.822$), $\eta_p{}^2 = 0.029$; S: $F_{2,60} = 0.320$, $p = 0.727$, $\eta_p{}^2 = 0.011$; ST: $F_{2,60} = 0.360$, $p = 0.699$, $\eta_p{}^2 = 0.01$, a significant effect of offset, T: $F_{2,60} = 5.413$, $p < 0.01$, $\eta_p{}^2 = 0.153$; S: $F_{2,60} = 3.751$, $p < 0.05$, $\eta_p{}^2 = 0.111$; ST: $F_{2,60} = 4.590$, $p < 0.05$, $\eta_p{}^2 = 0.133$, and a null onset × offset interaction, T: $F_{4,120} = 0.554$, $p = 0.696$,

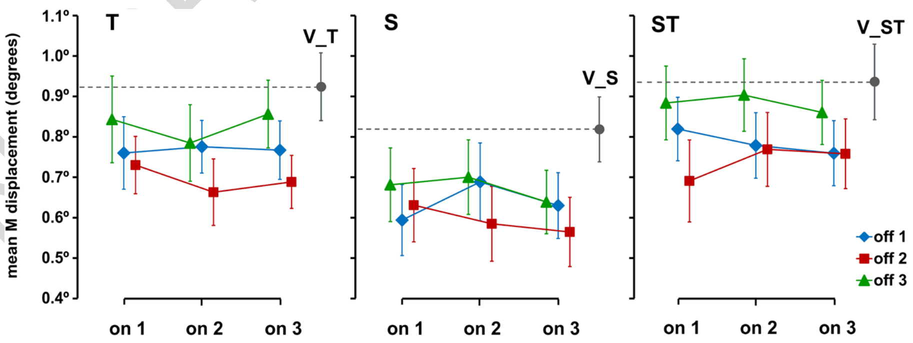

Figure 5. Onset × offset plots presented separately for each task (T, S and ST). The dashed line in each graph represents mean *M* displacement for interspersed visual-only presentations (V_T, V_S and V_ST). Mean *M* displacement is given on the ordinate. Error bars denote SEM.

UNCORRECTED PROOF

$\eta_p{}^2 = 0.018$; S: $F_{4,120} = 0.949$, $p = 0.438$, $\eta_p{}^2 = 0.031$, ST: $F_{4,120} = 0.940$, $p = 0.449$, $\eta_p{}^2 = 0.03$.

Across tasks, the factorial plots are located at different heights on the common ordinate, such that task S, at the center, presents the lower M displacement values and task ST, on the right, the higher ones, with task T in between. These differences between tasks were noted in Table 1 by the significant main effect of task ($p < 0.05$). No significant differences were found in subsequent pairwise comparisons (lowest observed $p = 0.064$, for S *versus* ST), but the relative decrease of M displacement in task S (middle level of the factor task) was indicated by a significant quadratic polynomial contrast, $F_{1,30} = 4.984$, $p < 0.05$, $\eta_p{}^2 = 0.142$.

Considering that all AV conditions in task S are synchronous, this might indicate that overall differences between tasks mostly depended on the presence (tasks T and ST) *versus* absence (task S) of temporal offset disparities, with a trend for lower M displacement in the latter case. However, it cannot be ruled out that spatial AV disparities in task S were responsible, in whole or in part, for the error decrease. Task ST, in which the same pattern of spatial disparities is replicated (in addition to temporal disparities), presents the highest M displacement values, suggesting that spatial disparities amplify somewhat the effects of temporal disparities, rather than contributing to error reduction. In any case, no definite conclusions seem possible based on the current design.

The dashed line in each graph of Fig56 represents M displacement obtained with visual-only intermixed presentations. These were the same across tasks but, as the graphics show, they present the same relative arrangement in height as their corresponding AV contexts (lowest M displacement for V_S, largest for V_ST). This raises the possibility of a contextual effect of the task on

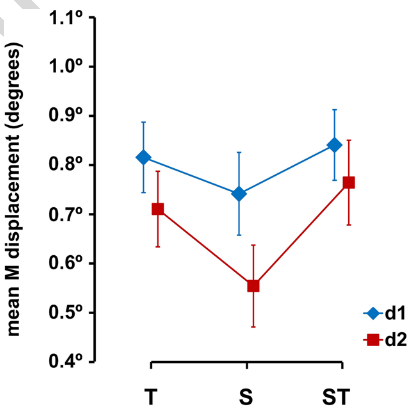

Figure 6. Factorial plots of task (T, S, ST) × distance (d1, d2) pooled across all the remaining factors. Mean *M* displacement is on the ordinate. Error bars denote standard errors of the mean.

UNCORRECTED PROOF

unimodal visual trials. However, a one-way repeated-measures ANOVA performed on visual intermixed conditions (V_T, V_S, V_ST) produced a non-significant result, $F_{1.417,41.093} = 1.804$ (ε correction $= 0.709$), $p = 0.185$, ${\eta_p}^2 = 0.059$.

The significant task × distance interaction in Table 1 ($p < 0.05$) is graphically illustrated in Fig. 6, showing a more pronounced decrease of M displacement in task S relative to the other two tasks in d2. One-way repeated-measures ANOVAs carried out separately for d1 and d2 revealed a null effect of task at d1, $F_{2,60} = 1.136$, $p = 0.264$, ${\eta_p}^2 = 0.043$, and a significant effect at d2, $F_{2,60} = 7.406$, $p = 0.001$, ${\eta_p}^2 = 0.198$. Follow-up pairwise comparisons disclosed further a significant difference between tasks S and ST at d2 ($p < 0.01$), while the comparison between tasks S and T yielded a $p = 0.06$.

Since d2 is associated with larger AV disparities, this could be interpreted in mainly two ways. Either the larger temporal disparities at T and ST produced stronger differences upward in M displacement (more error) regarding synchronous presentations (task S), or the larger spatial disparities in task S produced stronger differences downward (more error reduction) relative to the other tasks than smaller spatial disparities did in d1. The first interpretation appears more likely, supporting a primary role of temporal information in the differences between tasks. Nevertheless, as noticed, no definite conclusions can be made here regarding this issue.

3.5. Audiovisual and Unimodal Visual Conditions: Comparative Effects

A feature the graphs in Fig. 5 have in common is that all solid lines standing for AV conditions lay below the dashed line standing for V-only conditions. This is true without exception and for every level of onset. Hence, regardless of the presence of disparities, AV conditions invariably led to performance enhancement.

This error reduction was shown to be significant by paired t tests between V and AV trials (aggregated) performed separately for tasks T, S and ST: $t_{30} = 4.509$, $p < 0.001$; $t_{29} = 3.433$, $p < 0.01$; $t_{30} = 3.385$, $p < 0.01$. One participant was removed from the comparison in task S for yielding an extreme value (more than three times the interquartile range) in V trials. To further compare V trials with each of the AV offset conditions, separate one-way ANOVAs were performed in each task with V, off_1, off_2 and off_3 as levels of the within-subjects factor, to be followed by pairwise comparisons. In all tasks, M displacement at V was significantly higher than at off_2 and off_1. Task T: V ($M = 0.89°$, SD $= 0.45°$) > off_2 ($M = 0.67°$, SD $= 0.36°$), $p < 0.001$; V > off_1 ($M = 0.74°$, SD $= 0.38°$), $p < 0.05$. Task S: V ($M = 0.79°$, SD $= 0.43°$) > off_2 ($M = 0.57°$, SD $= 0.46°$), $p < 0.01$; V > off_1 ($M = 0.63°$, SD $= 0.43°$), $p < 0.05$. Task ST: V ($M = 0.90°$, SD $= 0.51°$) > off_2 ($M = 0.71°$, SD $= 0.47°$), $p = 0.001$; V > off_1 ($M = 0.76°$,

UNCORRECTED PROOF

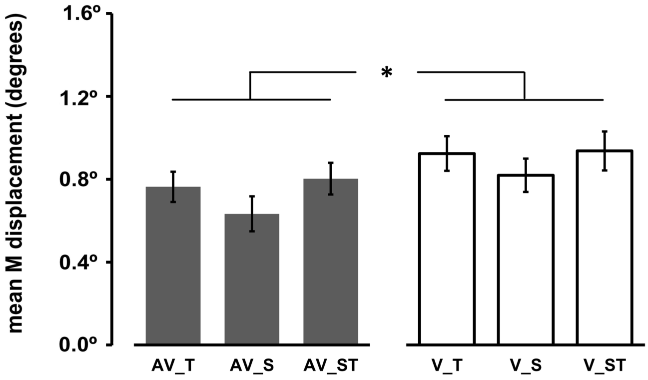

Figure 7. Comparative effects of AV and V presentations. Gray columns represent M displacement averaged over all AV presentations in each task (AV_T, AV_S and AV_ST). White columns correspond to mean *M* displacement for interspersed V conditions in each task (V_T, V_S and V_ST). Error bars stand for SEM.

SD = 0.40°), $p < 0.05$. Although consistent in all plots, the error reduction in AV off_3 relative to V did not reach statistical significance in any of the tasks (lowest observed $p = 0.145$).

Figure 7 provides a comparative overview of the effects of AV and unimodal V presentations across tasks (tasks denoted by letters after underscore). Larger forward errors in the block of V conditions (white columns) as compared to the block of AV conditions (gray columns) are apparent. The statistical difference observed in the graph between the two blocks (V and AV) was obtained in a repeated-measures multivariate analysis of variance (MANOVA) using type of condition as an independent grouping variable with two levels (V and AV), and tasks (T, S, ST) as three distinct measures of the effect (multiple dependent variables): Pillai's trace = 0.579, $F_{3,27} = 12{,}355$, $p < 0.001$.

In addition to comparing means between V and AV conditions, variances were also compared with the Pitman–Morgan test for correlated samples (one-tailed p values, unless stated otherwise). A significant reduction of variance in AV conditions relative to unimodal V conditions was found in tasks T and ST, but not in task S: V_T ($s^2 = 10.5°$) > AV_T ($s^2 = 8°$), $t_{29} = 1.748$, $p = 0.045$; V_ST ($s^2 = 13.2°$) > AV_ST ($s^2 = 8.7°$), $t_{29} = 2.774$, $p = 0.005$; V_S ($s^2 = 9.4°$) ≈ AV_S ($s^2 = 10.1°$), $t_{28} = 0.374$, $p = 0.356$. No significant differences in variance were found among tasks either in V (lower observed two-tailed $p = 0.264$) or in AV conditions (lower observed two-tailed $p = 0.382$). Summing up, results in this subsection show that bimodal AV conditions not only led to better average localization, but also (exception made to task S) to lower performance dispersion at the group level.

MSR (brill2x v1.36) msr1574.tex 15/04/2021 11:08 [research-article] p. 19/35

3.6. A Check on the Perceived Endpoints of Auditory Motion

Manipulation of the spatial offset of sounds was predicated on the assumption that perceived offsets would reflect both the true ordering of sound offset locations and their positions relative to visual offset. Otherwise, the finding that all AV conditions in tasks S and ST produce less forward error than the V conditions could result from the auditory spatial offset being always perceived as spatially prior to the visual offset, even when it was located ahead of it. Similarly, the invariant ordering of offset effects found across tasks (decreasing error reduction in the order off_2, off_1, off_3) might hypothetically result in tasks S and ST from spatial offset 2 being systematically (and wrongly) perceived as spatially prior to offset 1.

Such conjectures are limited at the outset by the fact that task T, where sounds were not spatially lateralized, exhibits the same pattern of results as the other tasks. Also, there is no obvious mechanism accounting for how off_2, for instance, could be systematically misperceived as spatially lagging behind off_1. In any case, the ancillary task in which participants localized the endpoint of unimodal auditory motion affords a direct test of those suppositions.

Figure 8 presents mean M displacement obtained for auditory motion (A) in the ancillary task (white bars) as well as, for comparison, M displacement in AV (gray bars) and intermixed V (black bars) trials in task S. The A stimuli consisted of the audio streams of AV presentations in task S where the audio and visual motion onsets coincided. For comparisons to be legitimate, thus, only that particular subset of AV presentations is considered in Fig. 8. The 0 value on the abscissa represents the veridical position of both visual offset (V_off) and auditory offset 2 (A_off_2), which always physically coincided. The values $-2°$ and $+2°$ correspond to the average true position of A_off_1 and A_off_3, respectively, in angular coordinates normalized to the visual offset location. M displacement is given by the bars in terms of deviations from the veridical value.

Two participants were removed for having results exceeding 1.5 times the interquartile range in at least one of the A_offset positions. The graph shows that unimodal auditory motion always produced a forward localization error, which was significantly different from zero at A_off_1 ($M = 0.82°$, $\pm 0.39°$ SEM), $t_{28} = 2.075$, $p < 0.05$, and A_off_3 ($M = 1.36°$, $\pm 0.32°$ SEM), $t_{28} = 4.042$, $p < 0.001$, but not at A_off_2 ($M = 0.6°$, $\pm 0.32°$ SEM), $t_{28} = 1.904$, $p = 0.067$. Mean error sizes were somewhat lower than those found in Getzmann *et al.* (2004), which ranged from 2° ($\pm 1.1°$ SEM) to 4° (± 1.2 SEM) for left–right movements, but overall comparable to the ones reported by Getzmann and Lewald (2009), of around 1.5° (± 0.5 SEM), and Feinkohl *et al.* (2014), of 0.88° and 1.26° (for target velocities of 25°/s

UNCORRECTED PROOF

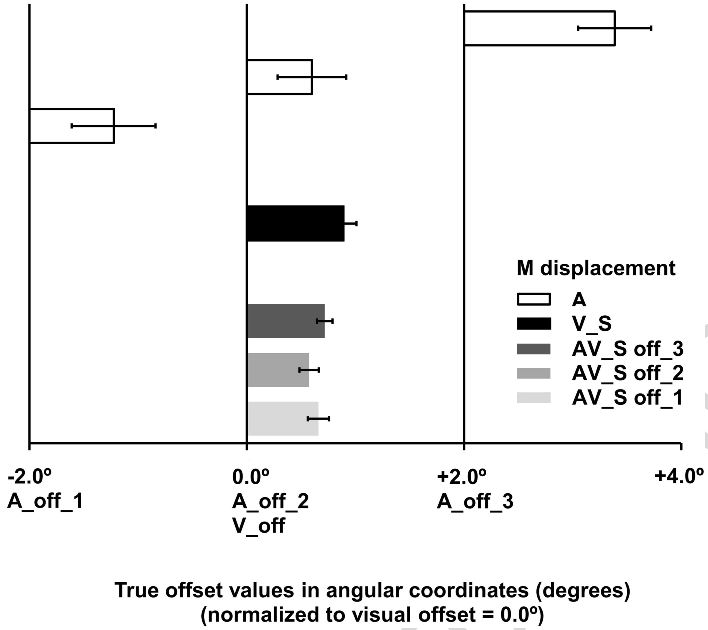

Figure 8. M displacement obtained for auditory motion (A) in the ancillary task (white bars), a subset of AV trials in task S with corresponding audio streams (gray bars) and intermixed visual-only trials in task S (black bars). The 0.0° value on the abscissa represents the veridical position of the physically coincident visual offset (V_off) and auditory offset 2 (A_off_2). The values −2.0° and +2.0° represent the average veridical position of A_off_1 and A_off_3, respectively, in angular coordinates normalized to the visual offset location. Average *M* displacement is given by the bars as a deviation from the true positions.

and 50°/s, respectively). Important differences remain between studies, concerning, e.g., response methods, how sound was presented, target velocities, vanishing points and travel durations. However, that our results are not too dissimilar from those in settings where sound spatialization has been taken a step further (Getzmann and collaborators use an array of loudspeakers to produce auditory motion; Feinkohl *et al*. use individualized HRTFs for presentation *via* headphones) can be seen as a favorable indication.

There were no significant differences in M displacement between the auditory offsets, $F_{2,56} = 2.803$, $p = 0.069$, $\eta_p{}^2 = 0.091$, nor between any of the A conditions and the intermixed V trials of task S (V_S: $M = 0.9°$; $\pm 0.12°$ SEM), $F_{3,84} = 1.905$, $p = 0.135$, $\eta_p{}^2 = 0.064$. This latter result agrees with Getzmann and Lewald's (2009) finding of similar forward displacements in both modalities (however, see Schmiedchen *et al*., 2013 for a different result). SEM values, represented by the error bars, are perceptibly larger in A conditions than in V_S conditions, and reach their minimum in AV conditions. Com-

UNCORRECTED PROOF

MSR (brill2x v1.36) msr1574.tex 15/04/2021 11:08 [research-article] p. 21/35

parisons with the Pitman–Morgan test revealed significant differences in variance between all A_off conditions and V_S ($p < 0.001$), as well as between A_off conditions and the corresponding AV conditions in task S ($p < 0.001$). No significant differences were found between V_S and any of the AV conditions, which all had lower associated variances: highest $t_{29} = 1.32$, $p = 0.197$, for AV offset 1.

Crucially for our purposes, A_off_1 was perceived as located behind both perceived A_off_2 and perceived V offset, and A_off_3 as located further ahead, as they should. Being physically coincident with V_S offset, A_off_2 was indeed perceived in the vicinity of perceived V_offset (0.29° behind it). A repeated-measures ANOVA performed over the physical coordinates indicated by the participants, with the three A_offsets and V_S as levels of a single within-subjects factor, yielded a significant result, $F_{3,84} = 66.327$, $p < 0.001$, $\eta_p{}^2 = 0.703$. Follow-up pairwise comparisons revealed significant differences for all pairs of A_offsets ($p < 0.001$) as well as for V_S *versus* A_off_1 ($p < 0.001$) and A_off_3 ($p < 0.001$), and a null result for V_S *versus* A_off_2 ($p = 1.00$). The relative locations of auditory spatial offsets, both with respect to each other and to visual offset, were thus correctly perceived. In addition, there does not seem to be any consistent way of deriving the perceived AV offsets (gray columns) from combining perceived A_offsets (white columns) and V_S offset. This rules out attraction toward the spatial location of the sound as a viable explanation of results in tasks S and ST.

4. Final Discussion and Conclusions

The preceding results support two main conclusions. First, across all tasks, onset disparities had no effects on the perceived endpoint of visual motion, while offset disparities yielded significant effects. This departs from Teramoto *et al.*'s (2010) finding that synchronous audiovisual onsets are needed for sound to modulate visual RM. It may be that Teramoto *et al.*'s (2010) result holds true for small AV asynchronies (<160 ms) and the sound attraction effects they reported. However, attraction effects were also obtained by Chien *et al.* (2013) with transient sounds, and under a similar range of asynchronies. Procedural differences could be causing these discrepancies, notably the use of a mouse-pointing response in both Chien *et al.* (2013) and the current study (as opposed to perceptual estimates). Specifically, processes in the sensorimotor branch of the visual system (Milner and Goodale, 1995), compounded in our study with oculomotor pursuit (Cui *et al.*, 2010; Jay and Sparks, 1984; Kubovy and Van Valkenburg, 2001; Zmigrod and Hommel, 2009), might have provided alternate sources of audiovisual bonding, dispensing with onset synchrony. Consistent with our own findings, Teramoto *et al.* (2010) reported a null effect of sound durations (extending up to around 1307 ms). This makes

UNCORRECTED PROOF

it unlikely that differences in stimulus durations explain divergences over the role of onset synchrony. A fair summary conclusion seems to be that our null finding places additional constraints on any notion that AV asynchronies at the onset inevitably preclude an impact of sound on visual RM.

The lack of effects of onset in task S has a methodological significance. Spatial disparities were conflated in this task with auditory travelled distance and speed, which both varied with spatial onset. Consequently, the null results for onset discard an influence of both these variables, allowing to unambiguously attribute the observed effects to spatial offset disparities. This indicates that correlating spatial AV disparities with auditory motion speed, rather than with motion duration (see Schmiedchen *et al.*, 2012), provides an effective way of dissociating temporal and spatial auditory cues by design.

The second conclusion concerns the modulatory effects of sounds. Though to varying degrees, performance enhancement occurred in all AV conditions. Importantly, in every task, the largest forward error reductions occurred when there were no offset disparities, not when sound offset was temporally or spatially prior to visual offset. The latter should have been the case if the perceived motion endpoint was being pulled toward sound offset. Finally, except for task S, AV conditions exhibited significant reductions in variance with respect to unimodal V conditions. Performance at the group level was thus not only more accurate but also more narrowly concentrated around the average localization response. This pattern of effects qualifies as bimodal enhancement, contrasting with the sound attraction bias reported in the comparison studies. Of note, while the goal assigned to participants was always to accurately locate the visual motion offset, only in our study did auditory information become instrumental to that purpose (i.e., functionally goal-relevant). This agrees with the expectation that using more salient disparities in fewer numbers could promote their relevance to the task.

More critically, we also posited that AV disparities might then tap into distinct multisensory processes. Teramoto *et al.* (2010) and Chien *et al.* (2013) accounted for their results in terms of multisensory binding, leading to emergent AV percepts. Schmiedchen *et al.* (2012) favored a pre-perceptual mechanism, yet likewise subjected to a narrow tolerance window for asynchronies. The prevailing role of tight temporal constraints is in all cases suggestive of stimulus-driven multisensory processes (Lamme and Roelfsema, 2000; Meredith and Stein, 1983; Meredith *et al.*, 1987). This makes in addition substantive sense, as the positional biases acted largely at odds with the task goal, appearing impervious to it.

Evidence consistent with an asynchrony tolerance window also turned up in our study, related to the equivalent benefits of audiovisual synchrony and of the sound-leading offset asynchrony in the shortest travelled distance d1 (−250 ms). This equivalence was replicated in tasks T and ST, where these

UNCORRECTED PROOF

conditions provided the largest benefits (in d1), and had no counterpart in spatial disparities (task S). Its asymmetry relative to 0 ms deviates from that of typical AV binding windows, usually tighter for sound-leading asynchronies (Conrey and Pisoni, 2006; Dixon and Spitz, 1980; Schmiedchen *et al.*, 2012; van Wassenhove *et al.*, 2007), but tallies with evidence that different mechanisms underlie tolerance for asynchronies depending on which modality has the lead (Cecere *et al.*, 2016). All considered, we take it as indicative of the involvement of early multisensory mechanisms, under a constraining temporal rule, at the 0 ms and −250 ms AV conditions. Targher *et al.* (2017) reported a similar equivalence between these exact same values, which they interpreted likewise.

The two main points calling for an explanation are then: why did synchronous conditions lead to the strongest benefits in the current study and no discernible enhancement in the comparison studies? And how did AV asynchronies far exceeding an effectiveness window not far from ±100 ms highlighted in the comparison studies still manage to produce beneficial effects in the current study? We tentatively suggest as a possible explanation to our finding of bimodal enhancement, valid for both points, the intervention of endogenous (goal-driven) attentional processing (Donohue *et al.*, 2015; Koelewijn *et al.*, 2010; Talsma *et al.*, 2009; Tang *et al.*, 2016; van Ee *et al.*, 2009). In this account, to release beneficial effects, task-relevant auditory timing information would need to be endogenously selected and made to cascade favorably into the visual processing of spatial location (Driver and Spence, 2000; Talsma *et al.*, 2010; Tang *et al.*, 2016). The promoted task-relevance of AV disparities in our study would have afforded a grip on endogenous attention not available in the comparison studies.

Cross-sensory cascading from one property to another has in turn been construed as a multisensory analog of object-based spread of attention (Donohue *et al.*, 2011; Driver and Spence, 2000; Fiebelkorn *et al.*, 2010, Molholm *et al.*, 2007; Tang *et al.*, 2016; Zimmer *et al.*, 2010a, b). In line with that, we accord a role in our tentative account to the distinction between stimulus-driven and representation-driven spread of attention (Fiebelkorn *et al.*, 2010; Molholm *et al.*, 2007). The former is based on percepts arising from bottom-up multisensory integration, which we surmised was the case at synchronous and −250 ms AV conditions. This could lead to not only stronger perceptual objecthood (Shomstein and Behrmann, 2008) but also exogenous capturing of attention by emerging AV percepts (Koelewijn *et al.*, 2010; Talsma *et al.*, 2010; Tang *et al.*, 2016), swiftly providing a basis for the spreading of endogenous attention. Both factors might thereby have underpinned the distinctive enhancement found at those conditions. For endogenous attention to take it from there it would have to selectively access auditory task-relevant features within the integrated percepts. This does not seem to be a particular problem, given the

UNCORRECTED PROOF

accepted prevalence of intermodal weak fusion (Clark and Yuille, 1990; Ernst and Bülthoff, 2004; Hillis *et al*., 2002; Landy *et al*., 1995), whereby unimodal information is not lost in multimodal objects. In contrast, exogenous attention alone would not be able to counteract the attraction bias in the comparison studies, and might even accentuate it.

Representation-driven spread of attention would conversely apply in the remaining conditions. Although conceived by Fiebelkorn *et al*. (2010) as mainly based on long-term multisensory associations between properties of well-known objects, different strands of evidence indicate that crossmodal correspondences learned during an experimental session may also originate coherent representations (Heron *et al*., 2012; Hidaka *et al*., 2015; Kim *et al*., 2009; Lewkowicz, 2003; Parise and Ernst, 2016; Parise *et al*., 2012; Seitz *et al*., 2007; Shams *et al*., 2011). The key feature here is that representations are no longer driven primarily by near co-occurrence (entailing specific temporal windows) but by supple detection-learning of patterned structures in the environment. This would account for the fact that even at longer asynchronies (e.g., ±500 ms) sounds still exert beneficial effects in our study, thought to a lesser degree, and that these effects can be obtained both pre- (sound-offset leading) and retro-actively (visual-offset leading).

Synchronous and −250 ms conditions were referred to stimulus-driven integration for empirical reasons already discussed. However, a −250 ms asynchrony still lies outside the temporal windows found in the comparison studies. This might be accounted as a contextual modulation of the boundaries of the integration window. Such temporal modulations are well documented and known to depend on several factors (Colonius and Diederich, 2010; Spence and Frings, 2020; Vroomen and Keetels, 2010), among which attention and task demands (Donohue *et al*., 2015; Mégevand *et al*., 2013). Attention to repeated AV contingencies could thus have also acted to reshape the limits of a suitable temporal window in the context of the task (Targher *et al*., 2017). This would be congenial, overall, to Colonius and Diederich's (2010) notion of a variable optimal time window of audiovisual integration.

Relatedly, the distinct effects of −250 ms and −500 ms asynchronies seem more adequately explained by the breath of an integration window than by conjecturable alternate mechanisms such as biphasic automatic cueing effects (e.g., inhibition of return and reflexive facilitation: Blurton *et al*., 2015) or temporal orienting (Rohenkohl *et al*., 2011). Experimental settings in which inhibition of return has been established (slower responses to targets presented at spatially cued peripheral locations than at uncued locations: Klein, 2000; Spence and Driver, 1998) appear too far removed from the current one to allow for useful transpositions. Moreover, the reported equivalence between synchronous and −250 ms AV conditions does not favor specific facilitation at the 250 ms cue–target interval. The framework of temporal orienting (as

UNCORRECTED PROOF

contrasted to spatial orienting) appears more fitting to the current setting, but it would lead in this case (equally valid temporal cues at −500 ms and −250 ms) to expect either similar or better results at −500 ms (Weinbach and Henik, 2012).

Some general objections to the preceding account might be based on nonspecific mechanisms. Enhancement might for instance have resulted from sound-leading asynchronies acting as nonspecific warning signals (see, e.g., Los and Van der Burg, 2013). Against that, in a variable foreperiod paradigm, reaction time (RT) advantages are known to increase with foreperiod length until up to −600 ms (Los and Van der Burg, 2013) and even −3 s (Niemi and Näätänen, 1981). AV asynchronies of −500 ms should then have been more beneficial than −250 ms asynchronies or the synchronous offset conditions. Also, sound offsets subsequent to visual offset, for which preparatory effects are excluded, still led to enhancement. Explaining this away as, e.g., a nonspecific response calibration effect would clash with the fact that visual-only trials did not benefit as much. Finally, nonspecific alerting has often been reported to increase speed at the expense of accuracy (Fernandez-Duque and Posner, 1997; Weinbach and Henik, 2012), yet accuracy gains were here the characteristic finding.

It could also be argued that AV enhancement would be most simply explained by a general response bias, consisting in responding more leftwards whenever a sound was presented. This would not explain the specific effects of offset disparities, or why no similar bias appeared in the comparison studies. Moreover, why a leftward bias should occur in the first place remains unclear. If it were a general correction mechanism, it should equally apply to visual-only conditions. If it were a consequence of the interruption of motion extrapolation (e.g., oculomotor overshooting: Kerzel, 2000, 2005) by abrupt sound offsets, it would still not account for the specific pattern of effects. Interestingly, a modified version of the latter conjecture, with extrapolation (e.g., Cooper and Munger, 1993; Freyd and Finke, 1984; Kerzel, 2003; Maus and Nijhawan, 2009) allayed by a more accurate timing of visual positional signals (crossmodally driven by sound), would fit both our suggested account and the whole of the results.

The effects of spatial AV disparities call for a cautious interpretation. Task S exhibited the most accurate responses, which could have simply followed from the fact that all conditions were synchronous. Given the current design, it cannot be established whether spatial auditory cues contributed any facilitatory effects over those already due to synchrony. Assessing that would require two additional tasks, one including only synchronous non-lateralized sounds, the other only spatiotemporally aligned conditions.

Misaligned AV offsets yielded significantly less error reduction than spatially aligned offsets (see Fig. 5). This establishes a direct (not temporally

UNCORRECTED PROOF

MSR (brill2x v1.36) msr1574.tex 15/04/2021 11:08 [research-article] p. 26/35

mediated) effect of auditory spatial cues on visual RM. It is also consistent with the finding that task ST, where spatial and temporal disparities covaried, showed the lowest amount of error reduction. As in the other tasks, the effects of spatial sound offsets were inconsistent with an attraction bias account. Results of the ancillary experiment (see Fig. 8), showing that the relative location of sound offsets was correctly perceived, additionally rule out explanations in terms of perceptual confusion over those locations.

Several aspects of sound presentation and spatialization may have limited the impact of spatial auditory information. Presentation *via* headphones is known to hamper out-of-the-head sound localization (Best *et al*., 2020) and may have countered audiovisual integration. The use of an average HRTF (S.-M. Kim and Choi, 2005; Mendonça *et al*., 2012), possible auditory coding ambiguities within the chosen bandwidth (Akeroyd, 2014; Macpherson and Middlebrooks, 2002), and broadband onset/offset splatter (Leshowitz and Wightman, 1972) may have also contributed to lessen the precision and ecological value of binaural cues. This notwithstanding, spatial offset disparities did produce significant effects on visual RM. Moreover, as visual-only conditions were invariably less accurate than AV conditions, accounting for the increased enhancement in task S by an increase in the relative weighting of visual information seems implausible. It may still be that the noticed shortcomings impacted the ability of spatial auditory cues to enhance performance. This is mostly a concern for future studies, as no conclusions regarding that issue can be drawn here, for reasons of design.

Differences in mean localization error between V-only conditions across tasks paralleled those between AV conditions (see Figs 5 and 7). While not statistically significant, they suggest a contextual effect of AV presentations upon unimodal visual trials. This could have relied on perceptual plasticity, criterion shifts or both. Subsequent improvement of unisensory visual processing following repeated exposure to AV associations has been documented (Shams *et al*., 2011) and suggested to be perceptual rather than cognitive (Kim *et al*., 2008). Anyhow, given the systematic advantage of AV conditions, this contextual effect should be regarded as distinct from the modulatory effect of sounds exerted within AV trials.

To conclude, this study makes a case for the ability of sounds to modulate visual RM through both temporal and spatial auditory cues in central vision. The whole of the findings, and how they compare to the literature, document a shift from goal-impervious attraction effects in prior studies to goal-oriented audiovisual enhancement in the present study. Accordingly, they support the hypothesis that manipulating the salience and/or reliability of AV disparities qualitatively changes the impact of sound on visual RM. This can be plausibly accounted for by a task-dependent interplay between multisensory processing and crossmodal attention (e.g., Busse *et al*., 2005; Fiebelkorn *et al*., 2010;

UNCORRECTED PROOF

Koelewijn *et al.*, 2010; Macaluso *et al.*, 2016; Staufenbiel *et al.*, 2011; Talsma *et al.*, 2010; Tang *et al.*, 2016). The role accorded to audiovisual contingencies not primarily driven by close temporal proximity aligns with evidence for the importance of learned associations in adaptive crossmodal perception (Heron *et al.*, 2010; Li *et al.*, 2015; Seitz *et al.*, 2006; Shams *et al.*, 2011), and a view of multisensory processing as committed to flexibly exploiting causal and statistical patterns in the environment (Ernst, 2007; Fiebelkorn *et al.*, 2010; Hidaka *et al.*, 2015; Kim *et al.*, 2008; Parise *et al.*, 2012; Seitz *et al.*, 2007). The study illustrates how, by treating the task context and demands as independent variables, seemingly conflicting findings can be possibly reconciled within a multiple-mechanisms view of crossmodal processing.

References

Akeroyd, M. A. (2014). An overview of the major phenomena of the localization of sound sources by normal-hearing, hearing-impaired, and aided listeners, *Trends Hear.* **18**, 2331216514560442. DOI:10.1177/2331216514560442.

Alais, D. and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration, *Curr. Biol.* **14**, 257–262. DOI:10.1016/j.cub.2004.01.029.

Battaglia, P. W., Jacobs, R. A. and Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization, *J. Opt. Soc. Am. A* **20**, 1391–1937. DOI:10.1364/JOSAA.20.001391.

Best, V., Baumgartner, R., Lavandier, M., Majdak, P. and Kopčo, N. (2020). Sound externalization: a review of recent research, *Trends Hear.* **24**, 2331216520948390. DOI:10.1177/2331216520948390.

Blurton, S. P., Greenlee, M. W. and Gondan, M. (2015). Cross-modal cueing in audiovisual spatial attention, *Atten. Percept. Psychophys.* **77**, 2356–2376. DOI:10.3758/s13414-015-0920-7.

Bruns, P. and Getzmann, S. (2008). Audiovisual influences on the perception of visual apparent motion: exploring the effect of a single sound, *Acta Psychol. (Amst).* **129**, 273–283. DOI:10.1016/j.actpsy.2008.08.002.

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. U. S. A.* **102**, 18751–18756. DOI:10.1073/pnas.0507704102.

Carlile, S. and Leung, J. (2016). The perception of auditory motion, *Trends Hear.* **20**, 1–19. DOI:10.1177/2331216516644254.

Cecere, R., Gross, J. and Thut, G. (2016). Behavioural evidence for separate mechanisms of audiovisual temporal binding as a function of leading sensory modality, *Eur. J. Neurosci.* **43**, 1561–1568. DOI:10.1111/ejn.13242.

Chandler, D. W. and Grantham, D. W. (1992). Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity, *J. Acoust. Soc. Am.* **91**, 1624–1636. DOI:10.1121/1.402443.

Chien, S., Ono, F. and Watanabe, K. (2013). A transient auditory signal shifts the perceived offset position of a moving visual object, *Front. Psychol.* **4**, 7. DOI:10.3389/fpsyg.2013.00070.

UNCORRECTED PROOF

Clark, J. J. and Yuille, A. L. (1990). *Data Fusion for Sensory Information Processing Systems*. Kluwer Academic Publishers, Boston, MA, USA.

Colonius, H. and Diederich, A. (2010). The optimal time window of visual-auditory integration: a reaction time analysis, *Front. Integr. Neurosci.* **4**, 11. DOI:10.3389/fnint.2010.00011.

Conrey, B. and Pisoni, D. B. (2006). Auditory-visual speech perception and synchrony detection for speech and nonspeech signals, *J. Acoust. Soc. Am.* **119**, 4065–4073. DOI:10.1121/1.2195091.

Cooper, L. A. and Munger, M. P. (1993). Extrapolating and remembering positions along cognitive trajectories: use and limitations of analogies to physical motion, in: *Spatial Representation: Problems in Philosophy and Psychology*, N. Eilan, R. A. McCarthy and B. Brewer (Eds), pp. 112–131. Blackwell Publishing, Malden, MA.

Cui, Q. N., Razavi, B., O'Neill, W. E. and Paige, G. D. (2010). Perception of auditory, visual, and egocentric spatial alignment adapts differently to changes in eye position, *J. Neurophysiol.* **103**, 1020–1035. DOI:10.1152/jn.00500.2009.

De Sá Teixeira, N. and Oliveira, A. M. (2011). Disambiguating the effects of target travelled distance and target vanishing point upon representational momentum, *J. Cogn. Psychol.* **23**, 650–658. DOI:10.1080/20445911.2011.557357.

De Sá Teixeira, N. and Oliveira, A. M. (2014). Spatial and foveal biases, not perceived mass or heaviness, explain the effect of target size on representational momentum and representational gravity, *J. Exp. Psychol. Learn. Mem. Cogn.* **40**, 1664–1679. DOI:10.1037/xlm0000011.

De Sá Teixeira, N. A., Hecht, H. and Oliveira, A. M. (2013). The representational dynamics of remembered projectile locations, *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 1690–1699. DOI:10.1037/a0031777.

Dick, M., Ullman, S. and Sagi, D. (1987). Parallel and serial processes in motion detection, *Science* **237**, 400–402. DOI:10.1126/science.3603025.

Dixon, N. F. and Spitz, L. (1980). The detection of auditory visual desynchrony, *Perception* **9**, 719–721. DOI:10.1068/p090719.

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. DOI:10.1523/JNEUROSCI.5298-10.2011.

Donohue, S. E., Green, J. J. and Woldorff, M. G. (2015). The effects of attention on the temporal integration of multisensory stimuli, *Front. Integr. Neurosci.* **9**, 32. DOI:10.3389/fnint.2015.00032.

Driver, J. and Spence, C. (2000). Multisensory perception: beyond modularity and convergence, *Curr. Biol.* **10**, R731–R735. DOI:10.1016/S0960-9822(00)00740-5.

Dupierrix, E., Alleysson, D., Ohlmann, T. and Chokron, S. (2008). Spatial bias induced by a non-conflictual task reveals the nature of space perception, *Brain Res.* **1214**, 127–135. DOI:10.1016/j.brainres.2008.01.021.

Eg, R. and Behne, D. M. (2015). Perceived synchrony for realistic and dynamic audiovisual events, *Front. Psychol.* **6**, 736. DOI:10.3389/fpsyg.2015.00736.

Eg, R., Griwodz, C., Halvorsen, P. and Behne, D. (2015). Audiovisual robustness: exploring perceptual tolerance to asynchrony and quality distortion, *Multimed. Tools Appl.* **74**, 345–365. DOI:10.1007/s11042-014-2136-6.

UNCORRECTED PROOF

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch, *J. Vis.* **7**, 7. DOI:10.1167/7.5.7.

Ernst, M. O. and Bülthoff, H. H. (2004). Merging the senses into a robust percept, *Trends Cogn. Sci.* **8**, 162–169. DOI:10.1016/j.tics.2004.02.002.

Feinkohl, A., Locke, S. M., Leung, J. and Carlile, S. (2014). The effect of velocity on auditory representational momentum, *J. Acoust. Soc. Am.* **136**, EL20–EL25. DOI:10.1121/1.4881318.

Fernandez-Duque, D. and Posner, M. I. (1997). Relating the mechanisms of orienting and alerting, *Neuropsychologia* **35**, 477–486. DOI:10.1016/S0028-3932(96)00103-0.

Fiebelkorn, I. C., Foxe, J. J. and Molholm, S. (2010). Dual mechanisms for the cross-sensory spread of attention: how much do learned associations matter?, *Cereb. Cortex* **20**, 109–120. DOI:10.1093/cercor/bhp083.

Frassinetti, F., Bolognini, N. and Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction, *Exp. Brain Res.* **147**, 332–343. DOI:10.1007/s00221-002-1262-y.

Freyd, J. J. (1987). Dynamic mental representations, *Psychol. Rev.* **94**, 427–438. DOI:10.1037/0033-295X.94.4.427.

Freyd, J. J. and Finke, R. A. (1984). Representational momentum, *J. Exp. Psychol. Learn. Mem. Cogn.* **10**, 126–132. DOI:10.1037/0278-7393.10.1.126.

García-Pérez, M. A. and Alcalá-Quintana, R. (2012). On the discrepant results in synchrony judgment and temporal-order judgment tasks: a quantitative model, *Psychon. Bull. Rev.* **19**, 820–846. DOI:10.3758/s13423-012-0278-y.

Getzmann, S. (2005). Shifting the onset of a moving sound source: a Fröhlich effect in spatial hearing, *Hear. Res.* **210**, 104–111. DOI:10.1016/j.heares.2005.08.003.

Getzmann, S. and Lewald, J. (2007). Localization of moving sound, *Percept. Psychophys.* **69**, 1022–1034. DOI:10.3758/BF03193940.

Getzmann, S. and Lewald, J. (2009). Constancy of target velocity as a critical factor in the emergence of auditory and visual representational momentum, *Exp. Brain Res.* **193**, 437–443. DOI:10.1007/s00221-008-1641-0.

Getzmann, S., Lewald, J. and Guski, R. (2004). Representational momentum in spatial hearing, *Perception* **33**, 591–599. DOI:10.1068/p5093.

Heron, J., Whitaker, D. and McGraw, P. V. (2004). Sensory uncertainty governs the extent of audio-visual interaction, *Vision Res.* **44**, 2875–2884. DOI:10.1016/j.visres.2004.07.001.

Heron, J., Roach, N. W., Whitaker, D. and Hanson, J. V. M. (2010). Attention regulates the plasticity of multisensory timing, *Eur. J. Neurosci.* **31**, 1755–1762. DOI:10.1111/j.1460-9568.2010.07194.x.

Heron, J., Roach, N. W., Hanson, J. V. M., McGraw, P. V. and Whitaker, D. (2012). Audiovisual time perception is spatially specific, *Exp. Brain Res.* **218**, 477–485. DOI:10.1007/s00221-012-3038-3.

Hidaka, S., Teramoto, W., Sugita, Y., Manaka, Y., Sakamoto, S. and Suzuki, Y. (2011). Auditory motion information drives visual motion perception, *PLoS ONE* **6**, e17499. DOI:10.1371/journal.pone.0017499.

Hidaka, S., Teramoto, W. and Sugita, Y. (2015). Spatiotemporal processing in crossmodal interactions for perception of the external world: a review, *Front. Integr. Neurosci.* **9**, 62. DOI:10.3389/fnint.2015.00062.

UNCORRECTED PROOF

Hillis, J. M., Ernst, M. O., Banks, M. S. and Landy, M. S. (2002). Combining sensory information: mandatory fusion within, but not between, senses, *Science* **298**, 1627–1630. DOI:10.1126/science.1075396.

Hommel, B. (2011). The Simon effect as tool and heuristic, *Acta Psychol. (Amst).* **136**, 189–202. DOI:10.1016/j.actpsy.2010.04.011.

Huang, P.-C., Yeh, S.-L., Chen, Y.-C. and Spence, C. (2011). Synchronous sounds enhance visual sensitivity without reducing target uncertainty, *Seeing Perceiv.* **24**, 623–638. DOI:10.1163/187847611X603765.

Hubbard, T. L. (1990). Cognitive representation of linear motion: possible direction and gravity effects in judged displacement, *Mem. Cognit.* **18**, 299–309. DOI:10.3758/BF03213883.

Hubbard, T. L. (1995a). Cognitive representation of motion: evidence for friction and gravity analogues, *J. Exp. Psychol. Learn. Mem. Cogn.* **21**, 241–254. DOI:10.1037/0278-7393.21.1.241.

Hubbard, T. L. (1995b). Environmental invariants in the representation of motion: implied dynamics and representational momentum, gravity, friction, and centripetal force, *Psychon. Bull. Rev.* **2**, 322–338. DOI:10.3758/BF03210971.

Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: a review of the findings, *Psychon. Bull. Rev.* **12**, 822–851. DOI:10.3758/BF03196775.

Hubbard, T. L. (2010). Approaches to representational momentum: theories and models, in: *Space and Time in Perception and Action*, R. Nijhawan and B. Khurana (Eds), pp. 338–365. Cambridge University Press, Cambridge, UK.

Hubbard, T. L. and Bharucha, J. J. (1988). Judged displacement in apparent vertical and horizontal motion, *Percept. Psychophys.* **44**, 211–221. DOI:10.3758/BF03206290.

Hubbard, T. L. and Lange, M. (2010). Prior probabilities and representational momentum: a signal detection analysis, *J. Vis.* **9**, 692. DOI:10.1167/9.8.692.

Hubbard, T. L. and Ruppel, S. E. (1999). Representational momentum and the landmark attraction effect, *Can. J. Exp. Psychol.* **53**, 242–256. DOI:10.1037/h0087313.

Ikumi, N. and Soto-Faraco, S. (2014). Selective attention modulates the direction of audio-visual temporal recalibration, *PLoS ONE* **9**, e99311. DOI:10.1371/journal.pone.0099311.

Jay, M. F. and Sparks, D. L. (1984). Auditory receptive fields in primate superior colliculus shift with changes in eye position, *Nature* **309**, 345–347. DOI:10.1038/309345a0.

Kawachi, Y. (2016). Visual mislocalization of moving objects in an audiovisual event, *PLoS ONE* **11**, e0154147. DOI:10.1371/journal.pone.0154147.

Kerzel, D. (2000). Eye movements and visible persistence explain the mislocalization of the final position of a moving target, *Vision Res.* **40**, 3703–3715. DOI:10.1016/S0042-6989(00)00226-1.

Kerzel, D. (2003). Mental extrapolation of target position is strongest with weak motion signals and motor responses, *Vision Res.* **43**, 2623–2635. DOI:10.1016/S0042-6989(03)00466-8.

Kerzel, D. (2005). Representational momentum beyond internalized physics: embodied mechanisms of anticipation cause errors in visual short-term memory, *Curr. Dir. Psychol. Sci.* **14**, 180–184. DOI:10.1111/j.0963-7214.2005.00360.x.

Kerzel, D. and Gegenfurtner, K. R. (2003). Neuronal processing delays are compensated in the sensorimotor branch of the visual system, *Curr. Biol.* **13**, 1975–1978. DOI:10.1016/j.cub.2003.10.054.

UNCORRECTED PROOF

Kim, S.-M. and Choi, W. (2005). On the externalization of virtual sound images in headphone reproduction: a Wiener filter approach, *J. Acoust. Soc. Am.* **117**, 3657–3665. DOI:10.1121/1.1921548.

Kim, R., Seitz, A. R. and Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning, *PLoS ONE* **3**, e1532. DOI:10.1371/journal.pone.0001532.

Kim, R., Seitz, A., Feenstra, H. and Shams, L. (2009). Testing assumptions of statistical learning: is it long-term and implicit?, *Neurosci. Lett.* **461**, 145–149. DOI:10.1016/j.neulet.2009.06.030.

Klein, R. M. (2000). Inhibition of return, *Trends Cogn. Sci.* **4**, 138–147. DOI:10.1016/S1364-6613(00)01452-2.

Koelewijn, T., Bronkhorst, A. and Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: a review of audiovisual studies, *Acta Psychol. (Amst).* **134**, 372–384. DOI:10.1016/j.actpsy.2010.03.010.

Krekelberg, B. and Lappe, M. (2001). Neuronal latencies and the position of moving objects, *Trends Neurosci.* **24**, 335–339. DOI:10.1016/S0166-2236(00)01795-1.

Kubovy, M. and Van Valkenburg, D. (2001). Auditory and visual objects, *Cognition* **80**, 97–126. DOI:10.1016/S0010-0277(00)00155-4.

Lamme, V. A. F. and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing, *Trends Neurosci.* **23**, 571–579. DOI:10.1016/S0166-2236(00)01657-X.

Landy, M. S., Maloney, L. T., Johnston, E. B. and Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion, *Vision Res.* **35**, 389–412. DOI:10.1016/0042-6989(94)00176-M.

Leshowitz, B. and Wightman, F. L. (1972). On the importance of considering the signal's frequency spectrum: some comments on Macmillans "Detection and recognition of increments and decrements in auditory intensity" experiment, *Percept. Psychophys.* **12**, 209–210. DOI:10.3758/BF03212872.

Lewkowicz, D. J. (2003). Learning and discrimination of audiovisual events in human infants: the hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues, *Dev. Psychol.* **39**, 795–804. DOI:10.1037/0012-1649.39.5.795.

Li, Q., Yang, H., Sun, F. and Wu, J. (2015). Spatiotemporal relationships among audiovisual stimuli modulate auditory facilitation of visual target discrimination, *Perception* **44**, 232–242. DOI:10.1068/p7846.

Lippert, M., Logothetis, N. K. and Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound, *Brain Res.* **1173**, 102–109. DOI:10.1016/j.brainres.2007.07.050.

Los, S. A. and Van der Burg, E. (2013). Sound speeds vision through preparation, not integration, *J. Exp. Psychol. Hum. Percept. Perform.* **39**, 1612–1624. DOI:10.1037/a0032183.

Macaluso, E., Noppeney, U., Talsma, D., Vercillo, T., Hartcher-O'Brien, J. and Adam, R. (2016). The curious incident of attention in multisensory integration: bottom-up *vs.* top-down, *Multisens. Res.* **29**, 557–583. DOI:10.1163/22134808-00002528.

Macpherson, E. A. and Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited, *J. Acoust. Soc. Am.* **111**, 2219–2236. DOI:10.1121/1.1471898.

UNCORRECTED PROOF

Matthews, N., Welch, L., Achtman, R., Fenton, R. and FitzGerald, B. (2016). Simultaneity and temporal order judgments exhibit distinct reaction times and training effects, *PLoS ONE* **11**, e0145926. DOI:10.1371/journal.pone.0145926.

Maus, G. W. and Nijhawan, R. (2009). Going, going, gone: localizing abrupt offsets of moving objects, *J. Exp. Psychol. Hum. Percept. Perform.* **35**, 611–626. DOI:10.1037/a0012317.

Mégevand, P., Molholm, S., Nayak, A. and Foxe, J. J. (2013). Recalibration of the multisensory temporal window of integration results from changing task demands, *PLoS ONE* **8**, e71608. DOI:10.1371/journal.pone.0071608.

Mendonça, C., Campos, G., Dias, P., Vieira, J., Ferreira, J. P. and Santos, J. A. (2012). On the improvement of localization accuracy with non-individualized HRTF-based sounds, *J. Audio Eng. Soc.* **60**, 821–830.

Meredith, M. A. and Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus, *Science* **221**, 389–391. DOI:10.1126/science.6867718.

Meredith, M. A., Nemitz, J. W. and Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors, *J. Neurosci.* **7**, 3215–3229. DOI:10.1523/JNEUROSCI.07-10-03215.1987.

Merz, S., Meyerhoff, H. S., Frings, C. and Spence, C. (2020). Representational momentum in vision and touch: visual motion information biases tactile spatial localization, *Atten. Percept. Psychophys.* **82**, 2618–2629. DOI:10.3758/s13414-020-01989-1.

Meyer, G. F., Wuerger, S. M., Röhrbein, F. and Zetzsche, C. (2005). Low-level integration of auditory and visual motion signals requires spatial co-localisation, *Exp. Brain Res.* **166**, 538–547. DOI:10.1007/s00221-005-2394-7.

Middlebrooks, J. C. and Green, D. M. (1991). Sound localization by human listeners, *Annu. Rev. Psychol.* **42**, 135–159. DOI:10.1146/annurev.ps.42.020191.001031.

Miller, J. D. and Wenzel, E. M. (2002). Recent developments in SLAB: a software based system for interactive spatial sound synthesis, in: *Proceedings of the 2002 International Conference on Auditory Display*, Kyoto, Japan, pp. 403–408.

Milner, A. D. and Goodale, M. A. (1995). *The Visual Brain in Action*. Oxford University Press, Oxford, UK.

Molholm, S., Martinez, A., Shpaner, M. and Foxe, J. J. (2007). Object-based attention is multisensory: co-activation of an object's representations in ignored sensory modalities, *Eur. J. Neurosci.* **26**, 499–509. DOI:10.1111/j.1460-9568.2007.05668.x.

Morein-Zamir, S. and Meiran, N. (2003). Individual stopping times and cognitive control: converging evidence for the stop signal task from a continuous tracking paradigm, *Q. J. Exp. Psychol. A* **56**, 469–489. DOI:10.1080/02724980244000495.

Morgan, W. A. (1939). A test for the significance of the difference between the two variances in a sample from a normal bivariate population, *Biometrika* **31**, 13–19. DOI:10.2307/2334972.

Munger, M. P., Dellinger, M. C., Lloyd, T. G., Johnson-Reid, K., Tonelli, N. J., Wolf, K. and Scott, J. M. (2006). Representational momentum in scenes: learning spatial layout, *Mem. Cognit.* **34**, 1557–1568. DOI:10.3758/BF03195919.

Niemi, P. and Näätänen, R. (1981). Foreperiod and simple reaction time, *Psychol. Bull.* **89**, 133–162. DOI:10.1037/0033-2909.89.1.133.

Noesselt, T., Bergmann, D., Hake, M., Heinze, H.-J. and Fendrich, R. (2008). Sound increases the saliency of visual events, *Brain Res.* **1220**, 157–163. DOI:10.1016/j.brainres.2007.12.060.

UNCORRECTED PROOF

Parise, C. V. and Ernst, M. O. (2016). Correlation detection as a general mechanism for multisensory integration, *Nat. Commun.* **7**, 11543. DOI:10.1038/ncomms11543.

Parise, C. V., Spence, C. and Ernst, M. O. (2012). When correlation implies causation in multisensory integration, *Curr. Biol.* **22**, 46–49. DOI:10.1016/j.cub.2011.11.039.

Perrott, D. R. and Musicant, A. D. (1977). Minimum auditory movement angle: binaural localization of moving sound sources, *J. Acoust. Soc. Am.* **62**, 1463–1466. DOI:10.1121/1.381675.

Perrott, D. R. and Saberi, K. (1990). Minimum audible angle thresholds for sources varying in both elevation and azimuth, *J. Acoust. Soc. Am.* **87**, 1728–1731. DOI:10.1121/1.399421.

Perrott, D. R., Costantino, B. and Cisneros, J. (1993). Auditory and visual localization performance in a sequential discrimination task, *J. Acoust. Soc. Am.* **93**, 2134–2138. DOI:10.1121/1.406675.

Pitman, E. J. G. (1939). A note on normal correlation, *Biometrika* **31**, 9–12. DOI:10.2307/2334971.

Risoud, M., Hanson, J.-N., Gauvrit, F., Renard, C., Lemesre, P.-E., Bonne, N.-X. and Vincent, C. (2018). Sound source localization, *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* **135**, 259–264. DOI:10.1016/j.anorl.2018.04.009.

Rohenkohl, G., Coull, J. T. and Nobre, A. C. (2011). Behavioural dissociation between exogenous and endogenous temporal orienting of attention, *PLoS ONE* **6**, e14620. DOI:10.1371/journal.pone.0014620.

Schmiedchen, K., Freigang, C., Nitsche, I. and Rübsamen, R. (2012). Crossmodal interactions and multisensory integration in the perception of audio-visual motion — a free-field study, *Brain Res.* **1466**, 99–111. DOI:10.1016/j.brainres.2012.05.015.

Schmiedchen, K., Freigang, C., Rübsamen, R. and Richter, N. (2013). A comparison of visual and auditory representational momentum in spatial tasks, *Atten. Percept. Psychophys.* **75**, 1507–1519. DOI:10.3758/s13414-013-0495-0.

Seitz, A. R., Kim, R. and Shams, L. (2006). Sound facilitates visual learning, *Curr. Biol.* **16**, 1422–1427. DOI:10.1016/j.cub.2006.05.048.

Seitz, A. R., Kim, R., van Wassenhove, V. and Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations, *Perception* **36**, 1445–1453. DOI:10.1068/p5843.

Shams, L., Wozny, D. R., Kim, R. and Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing, *Front. Psychol.* **2**, 264. DOI:10.3389/fpsyg.2011.00264.

Shomstein, S. and Behrmann, M. (2008). Object-based attention: strength of object representation and attentional guidance, *Percept. Psychophys.* **70**, 132–144. DOI:10.3758/PP.70.1.132.

Snedecor, G. W. and Cochran, W. G. (1989). *Statistical Methods*, 8th edn. Iowa State University Press, Ames, IA, USA.

Spence, C. and Driver, J. (1998). Auditory and audiovisual inhibition of return, *Percept. Psychophys* **60**, 125–139. DOI:10.3758/BF03211923.

Spence, C. and Frings, C. (2020). Multisensory feature integration in (and out) of the focus of spatial attention, *Atten. Percept. Psychophys.* **82**, 363–376. DOI:10.3758/s13414-019-01813-5.

UNCORRECTED PROOF

Staufenbiel, S. M., van der Lubbe, R. H. J. and Talsma, D. (2011). Spatially uninformative sounds increase sensitivity for visual motion change, *Exp. Brain Res.* **213**, 457–464. DOI:10.1007/s00221-011-2797-6.

Stein, B. E., London, N., Wilkinson, L. K. and Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis, *J. Cogn. Neurosci.* **8**, 497–506. DOI:10.1162/jocn.1996.8.6.497.

Stork, S. and Müsseler, J. (2004). Perceived localizations and eye movements with action-generated and computer-generated vanishing points of moving stimuli, *Vis. Cogn.* **11**, 299–314. DOI:10.1080/13506280344000365.

Talsma, D., Senkowski, D. and Woldorff, M. G. (2009). Intermodal attention affects the processing of the temporal alignment of audiovisual stimuli, *Exp. Brain Res.* **198**, 313–328. DOI:10.1007/s00221-009-1858-6.

Talsma, D., Senkowski, D., Soto-Faraco, S. and Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration, *Trends Cogn. Sci.* **14**, 400–410. DOI:10.1016/j.tics.2010.06.008.

Tang, X., Wu, J. and Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention, *Neurosci. Biobehav. Rev.* **61**, 208–224. DOI:10.1016/j.neubiorev.2015.11.002.

Targher, S., Micciolo, R., Occelli, V. and Zampini, M. (2017). The role of temporal disparity on audiovisual integration in low-vision individuals, *Perception* **46**, 1356–1370. DOI:10.1177/0301006617720124.

Teramoto, W., Hidaka, S., Gyoba, J. and Suzuki, Y. (2010). Auditory temporal cues can modulate visual representational momentum, *Atten. Percept. Psychophys.* **72**, 2215–2226. DOI:10.3758/BF03196696.

Teramoto, W., Hidaka, S., Sugita, Y., Sakamoto, S., Gyoba, J., Iwaya, Y. and Suzuki, Y. (2012). Sounds can alter the perceived direction of a moving visual object, *J. Vis.* **12**, 11. DOI:10.1167/12.3.11.

Teramoto, W., Hidaka, S. and Sugita, Y. (2018). Auditory biases in visual motion perception, in: *Spatial Biases in Perception and Cognition*, T. L. Hubbard (Ed.), pp. 167–179. Cambridge University Press, Cambridge, UK.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W. and Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search, *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 1053–1065. DOI:10.1037/0096-1523.34.5.1053.

van Ee, R., van Boxtel, J. J. A., Parker, A. L. and Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection, *J. Neurosci.* **29**, 11641–11649. DOI:10.1523/JNEUROSCI.0873-09.2009.

van Eijk, R. L. J., Kohlrausch, A., Juola, J. F. and van de Par, S. (2008). Audiovisual synchrony and temporal order judgments: effects of experimental method and stimulus type, *Percept. Psychophys.* **70**, 955–968. DOI:10.3758/PP.70.6.955.

van Wassenhove, V., Grant, K. W. and Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception, *Neuropsychologia* **45**, 598–607. DOI:10.1016/j.neuropsychologia.2006.01.001.

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. DOI:10.1037/0096-1523.26.5.1583.

UNCORRECTED PROOF

Vroomen, J. and de Gelder, B. (2004). Temporal ventriloquism: sound modulates the flash-lag effect, *J. Exp. Psychol. Hum. Percept. Perform.* **30**, 513–518. DOI:10.1037/0096-1523.30.3.513.

Vroomen, J. and Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review, *Atten. Percept. Psychophys.* **72**, 871–884. DOI:10.3758/APP.72.4.871.

Watanabe, K. and Shimojo, S. (2001). When sound affects vision: effects of auditory grouping on visual motion perception, *Psychol. Sci.* **12**, 109–116. DOI:10.1111/1467-9280.00319.

Weinbach, N. and Henik, A. (2012). Temporal orienting and alerting — the same or different?, *Front. Psychol.* **3**, 236. DOI:10.3389/fpsyg.2012.00236.

Werner, S. and Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization, *J. Neurosci.* **30**, 2662–2675. DOI:10.1523/JNEUROSCI.5091-09.2010.

Whitney, D. (2002). The influence of visual motion on perceived position, *Trends Cogn. Sci.* **6**, 211–216. DOI:10.1016/S1364-6613(02)01887-9.

Yost, W. A. and Zhong, X. (2014). Sound source localization identification accuracy: bandwidth dependencies, *J. Acoust. Soc. Am.* **136**, 2737–2746. DOI:10.1121/1.4898045.

Yost, W. A., Loiselle, L., Dorman, M., Burns, J. and Brown, C. A. (2013). Sound source localization of filtered noises by listeners with normal hearing: a statistical analysis, *J. Acoust. Soc. Am.* **133**, 2876–2882. DOI:10.1121/1.4799803.

Zimmer, U., Itthipanyanan, S., Grent-'t-Jong, T. and Woldorff, M. G. (2010a). The electrophysiological time course of the interaction of stimulus conflict and the multisensory spread of attention, *Eur. J. Neurosci.* **31**, 1744–1754. DOI:10.1111/j.1460-9568.2010.07229.x.

Zimmer, U., Roberts, K. C., Harshbarger, T. B. and Woldorff, M. G. (2010b). Multisensory conflict modulates the spread of visual attention across a multisensory object, *Neuroimage* **52**, 606–616. DOI:10.1016/j.neuroimage.2010.04.245.

Zmigrod, S. and Hommel, B. (2009). Auditory event files: integrating auditory perception and action planning, *Atten. Percept. Psychophys.* **71**, 352–362. DOI:10.3758/APP.71.2.352.

UNCORRECTED PROOF