Attentional modulations of audiovisual interactions in apparent motion: Temporal ventriloquism effects on perceived visual speed

Aysun Duyar [1,2,3] · Andrea Pavan [4] · Hulusi Kafaligonul [1,5]

Abstract

The timing of brief stationary sounds has been shown to alter different aspects of visual motion, such as speed estimation. These effects of auditory timing have been explained by temporal ventriloquism and auditory dominance over visual information in the temporal domain. Although previous studies provide unprecedented evidence for the multisensory nature of speed estimation, how attention is involved in these audiovisual interactions remains unclear. Here, we aimed to understand the effects of spatial attention on these audiovisual interactions in time. We utilized a set of audiovisual stimuli that elicit temporal ventriloquism in visual apparent motion and asked participants to perform a speed comparison task. We manipulated attention either in the visual or auditory domain and systematically changed the number of moving objects in the visual field. When attention was diverted to a stationary object in the visual field via a secondary task, the temporal ventriloquism effects on perceived speed decreased. On the other hand, focusing attention on the auditory stimuli facilitated these effects consistently across different difficulty levels of secondary auditory task. Moreover, the effects of auditory timing on perceived speed did not change with the number of moving objects and existed in all the experimental conditions. Taken together, our findings revealed differential effects of allocating attentional resources in the visual and auditory domains. These behavioral results also demonstrate that reliable temporal ventriloquism effects on visual motion can be induced even in the presence of multiple moving objects in the visual field and under different perceptual load conditions.

Keywords Motion · Attention · Speed perception · Multisensory · Audiovisual interactions · Temporal processing

Introduction

Motion perception is an important aspect of our daily experience. To perform proper actions and interact with a dynamic environment, humans (and many other species) precisely estimate the direction and speed of moving objects. Accordingly, motion processing has become an extensively investigated visual feature (Burr & Thompson, 2011; Kolers, 1972; Nakayama, 1985; Nishida, 2011). In studies investigating motion perception, the manipulations have been mainly based on visual stimulation and hence restricted to the visual modality. On the other hand, multisensory research ushered a new perspective of motion perception, wherein the information provided by other modalities (e.g., audition) is also involved in computations underlying motion perception (Soto-Faraco et al., 2003; Soto-Faraco & Väljamäe, 2012). To date, various audiovisual paradigms have been developed to demonstrate the multisensory nature of motion processing. Of particular relevance to the current study, the timing of brief static sounds (e.g., clicks) can alter apparent motion perception (Getzmann, 2007; Shi et al., 2010). Specifically, the time interval between static clicks has been found to modulate perceived direction, speed, and sensitivity to visual apparent motion (Freeman & Driver, 2008; Kafaligonul & Stoner, 2010, 2012; Ogulmus et al., 2018).

In these studies, the experimental design is typically based on two-frame apparent motion. Two concurrent brief sounds (e.g., clicks) have been used for auditory stimulation, and the

✉ Hulusi Kafaligonul
  hulusi@bilkent.edu.tr

1   National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, Turkey

2   Department of Psychology, Bilkent University, Ankara, Turkey

3   Department of Psychology, New York University, New York, NY, USA

4   Department of Psychology, University of Bologna, Viale Berti Pichat, 5, 40127 Bologna, Italy

5   Interdisciplinary Neuroscience Program, Aysel Sabuncu Brain Research Center, Bilkent University, 06800 Ankara, Turkey

time interval between them is systematically changed. The auditory time interval of these static sounds has been shown to modulate different aspects of motion perception. For example, previous research indicated that auditory time intervals can alter the perceived speed of two-frame apparent motion (Kafaligonul & Stoner, 2010; Ogulmus et al., 2018). The apparent motion with a short auditory time interval is perceived to move faster than the one with a long time interval, although apparent motions are the same in terms of visual stimulation. These effects of auditory timing on apparent motion percept have been interpreted as a consequence of a well-known phenomenon called "temporal ventriloquism." In general, temporal ventriloquism refers to the ability of brief sounds to drive the perceived timing of brief visual events when these stimuli are presented at different times (Fendrich & Corballis, 2001; Morein-Zamir et al., 2003; Recanzone, 2003). This illusion makes adaptive sense given the auditory system's superior temporal resolution, and such dominance has been mostly described as brief sounds affecting (e.g., capturing) visual events in time (Burr et al., 2009; Vroomen & Keetels, 2010; Welch & Warren, 1980). In the case of two-frame apparent motion paradigms, the static clicks may similarly drive the timing of visual motion frames (or the time interval between them). Hence, a decrease or an increase in the perceived time interval between the two motion frames may lead to faster and slower motion percepts, respectively.

The effects of auditory time interval on apparent motion provide important evidence that audiovisual interactions in the temporal domain play a critical role in motion perception. There is also neurophysiological evidence that auditory timing can affect the amplitude of evoked activities at both early and later stages of motion processing (Kaya et al., 2017; Kaya & Kafaligonul, 2019). These findings suggested that the effects of auditory time intervals on motion perception may be the outcome of a dynamic interplay between different cortical regions. An important question to address is how attention is involved in these interactions at different stages of cortical processing. Attention allows prioritization of relevant information for further processing according to context and task demands. The role of attention is complicated and context-dependent in crossmodal interactions. An emerging notion suggests that multisensory processing and attention interact in a complex, multifaceted manner. In agreement with this perspective, mounting evidence suggests that attention can take place at different levels of multisensory processing (Teder-Sälejärvi et al., 1999). Furthermore, the bottom-up (stimulus-driven) and top-down (goal-driven) attention may have differential effects at distinct stages of processing (Koelewijn et al., 2010; Macaluso et al., 2016; Talsma et al., 2010). Spatial attention can affect processing across sensory modalities, such that the processing of irrelevant visual information is enhanced in the attended (auditory) location and vice versa (Spence & Driver, 1996). In particular, attentional

allocation enhances perception across sensory modalities in motion perception (e.g., Beer & Röder, 2004a, 2004b). Attentional demands increase with additional tasks and/or with the task difficulty, which results in increased perceptual load. Perceptual load can influence audiovisual interactions in space, as well as the speed of audiovisual feature binding (e.g., Alsius et al., 2005; Eramudugolla et al., 2011; Evans, 2020).

Freeman and Driver (2008) investigated whether this form of audiovisual motion illusion (i.e., temporal ventriloquism effects on apparent motion) may be achieved simply by focusing attention on specific visual intervals. The auditory clicks may conceivably capture attention, potentially making some intervals between apparent motion frames more salient than others and affecting motion perception without changing the perceived visual timing. Their behavioral findings rejected this hypothesis based on the attention-capture account. Moreover, Kafaligonul and Stoner (2012) aimed to understand the involvement of attention-based motion system. They found that click timing can affect visual motion processing even when attentional tracking is ruled out (i.e., without the involvement of higher-order attentional and/or position tracking mechanisms). Therefore, these previous studies suggest that attention may not be required for this audiovisual temporal illusion to occur, highlighting the automatic nature of audiovisual interactions. Nevertheless, attention can have a modulatory influence on these audiovisual interactions in time and little is known about such modulatory role. This is mainly because visual apparent motion and auditory clicks were primary and secondary task-irrelevant stimuli in previous research, respectively. In other words, observers performed a perceptual task on visual motion while passively listening to the static clicks. Accordingly, the observers focused their attention on visual motion, and there was no systematic manipulation of attention either in the visual field or across modalities. On the other hand, such manipulations of attention have important implications for daily life situations.

In everyday life, the stimulation of the external environment is complex, and we are frequently exposed to more than one moving object in the visual field. Furthermore, the sensory relevance and attentional demands constantly change. Using complex stimulus configurations, previous research investigated the roles of feature similarity and crossmodal correspondence in temporal ventriloquism (Boyce, Lindsay, et al., 2020a; see also Chen et al., 2018). Although previous findings revealed significant effects of similarity, they also indicated that the featural differences did not abolish temporal ventriloquism (Boyce, Whiteford, et al., 2020b; Klimova et al., 2017). This applies to the number of stimuli in the visual and auditory domains. Against the original descriptions (Morein-Zamir et al., 2003), an equal number of auditory and visual stimuli (e.g., the number of visual objects and clicks) may not be necessary to elicit temporal ventriloquism effects on the perception of apparent motion (Getzmann,

2007; Ogulmus et al., 2018). Besides having important implications for audiovisual binding in the temporal domain (see Experiment 1), these results pave the way to investigate the role of spatial attention and to manipulate sensory relevance and attentional demands. Within the context of temporal ventriloquism effects on perceived speed, there is still no systematic research on the number of visual stimuli and the role of spatial attention in these audiovisual interactions. An important question is whether the auditory time interval can alter the perception of more than one moving object and when the attention is distributed within the visual field. In the present study, we first aimed to address this question by investigating the effects of auditory time interval on speed perception. We systematically manipulated the number of concurrent moving objects in the visual field under different attention conditions. Additionally, we included a secondary perceptual task on the visual events (i.e., a dual-task paradigm) to assess the allocation of attentional resources. We next asked whether focusing attentional resources on the auditory click would modulate these audiovisual interactions in time. In this part of the study, we introduced a secondary task on the location of static clicks and systematically manipulated the secondary task difficulty by shifting the position of the sound source, which also allowed us to examine whether the possible modulations due to perceptual load on the auditory stimulation depend on task difficulty.

Experiment 1

Using a visual search (i.e., pip and pop) paradigm, previous research revealed that audiovisual integration decreases drastically with more than one static object in the visual field (Olivers et al., 2016; Van der Burg et al., 2013). According to these findings, the number of visual events that may be linked to a single auditory event is limited. On the other hand, behavioral studies combining temporal ventriloquism and apparent motion indicated that auditory time intervals can affect more than one moving object (e.g., Ogulmus et al., 2018). These findings suggest that the timing of a single auditory click may drive the timing of more than one object presented in each motion frame, because two-frame apparent motion and two concurrent clicks were typically used in previous research, and the effects of temporal ventriloquism have been mostly described as each click affecting the perceived timing of each apparent motion frame (or the time interval demarcated by these frames; Chen & Vroomen, 2013). However, there is still no systematic investigation on testing the limits of these audiovisual interactions in terms of the number of moving objects in the visual field. Therefore, in the first experiment, we examined auditory time interval effects on perceived speed by systematically manipulating the number of moving objects and spatial attention in the visual

field. Based on the hypothesis that there is a limited capacity of binding auditory and visual events, we expected to have an increase in the amount of temporal ventriloquism effects on perceived visual speed when observers attended to a single moving object in the visual field.

Moreover, dual-task paradigms (i.e., having a secondary task) have been used to manipulate attentional resources in multisensory paradigms. Previous work showed that attentional demands modulate audiovisual processing and binding (e.g., Alsius et al., 2005; Mozolic et al., 2008; Ren et al., 2020; Ren et al., 2021). Using a secondary task in the visual domain, these studies indicated that audiovisual interactions were greatly reduced when participants concurrently performed an unrelated visual task. Accordingly, we also assessed whether the allocation of attentional resources in the visual domain alters the amount of auditory time interval effects on perceived speed by introducing a secondary task on the fixation target. Based on the previous research on different audiovisual paradigms, we hypothesized that diverting attention away from moving stimuli would decrease the binding and hence audiovisual interactions in time.

Methods

Participants

Twelve participants (age range: 21–29 years) completed all the training and main experimental sessions. All participants had normal or corrected-to-normal vision and normal hearing. None had a history of neurological disorders by self-report. Before their participation, they were informed about experimental procedures and signed a consent form. The sample size was determined based on our previous behavioral studies examining the effects of auditory time interval on perceived visual speed (Kafaligonul & Stoner, 2010; Ogulmus et al., 2018; see also the behavioral study reported in Kaya et al., 2017). In particular, Ogulmus et al. (2018) used a design based on comparing two consecutive apparent motions with different auditory time intervals. All the sample sizes reported in the present study were also commensurate with the original research by Van der Burg et al. (2013) investigating the capacity of audiovisual binding. All procedures were in accordance with the Declaration of Helsinki (World Medical Association, 2013) and approved by the local Ethics Committee of Bilkent University.

Apparatus

We used MATLAB (The MathWorks, Natick, MA, USA) with the Psychtoolbox 3.0 extension (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) to control stimulation, experimental design, and data acquisition. The visual stimuli were displayed on a 20-inch CRT screen (1,280 × 1,024–pixel

resolution, 100-Hz refresh rate) at a viewing distance of 57 cm. The display was gamma-corrected using a SpectroCAL (Cambridge Research Systems, Rochester, Kent, UK) photometer. The auditory stimuli were emitted by two-channel speakers positioned next to the display on each side. The center of speakers (i.e., the horizontal midpoint between the two speakers) was vertically aligned with the display and 57 cm away from the participants. The sound pressure level (SPL) was regularly measured with a sound-level meter (SL-4010 Lutron, Lutron Electronics, Taipei, TW). A chin rest was used to stabilize the head position and constrain movements. The experiments were performed in a dimly lit and sound-attenuated testing room. Except for a speaker change in Experiment 4, the same apparatus and testing room were used in all the experiments.

Stimuli and procedure

The design was based on comparing the speed of two consecutive apparent motions moving in the same direction (Kafaligonul & Stoner, 2010; Ogulmus et al., 2018). A small square (0.5° length, 108 cd/m^2) at the center of the display (0.56 cd/m^2 background luminance) served as a fixation marker. Each apparent motion consisted of two motion frames (Fig. 1a). In each motion frame, an equal number of objects (2, 4, or 8 objects) were presented on an imaginary circle (inner circle radius: 2.15°, outer circle radius: 3.85°) around the fixation. The shape of each object was pseudorandomly assigned to a circle (0.6° diameter, 54.5 cd/m^2) or a square (0.6° length, 54.5 cd/m^2). When there were two objects, the stimuli were positioned on the left and right side of the fixation. Therefore, the resulting movement was horizontal, and there was a 180° angle between the motion directions. For the 4 and 8 object presentations, the positions of objects were equally spaced in each frame to have 90° and 45° angles between neighboring motion directions, respectively (Fig. 1b). Apparent motions were generated by presenting each frame for 50 ms and having a 100-ms blank interval between them (ISI$_v$, interstimulus interval). During the blank interval, there was only the fixation at the center of the display (Fig. 1a). Based on the overall motion direction (outwards or inwards) during a trial, the motion frame in which objects were positioned either on the inner circle or the outer circle was presented first. A pair of static clicks was also introduced during the presentation of each two-frame apparent motion. Each click had a duration of 20 ms (rectangular windowed 480-Hz sine-wave carrier, 44.1-kHz sampling rate), and the SPL was 78 dB. The pair of clicks was introduced with a time interval (ISI$_a$) and temporally centered with respect to the pair of motion frames.

For each trial, the number of objects in an apparent motion frame was pseudorandomly selected from the three conditions (2, 4, or 8 objects). The two-frame apparent motion stimuli were shown twice. The interval between each consecutive presentation was 700 ms (i.e., the ISI between the first and second apparent motion presentation, see Fig. 1a for the timeline). Each apparent motion was the same, but the auditory time interval between the concurrent sounds was different. For one of the apparent motion presentations, the time interval between static clicks was shorter than the visual time interval between the two motion frames (short ISI$_a$ = 20 ms). For the other one, the auditory time interval was longer than the visual time interval (long ISI$_a$ = 240 ms). The order of auditory time intervals (short vs. long) was randomized across trials. The timeline of events, including auditory time intervals were based on previous studies (Kafaligonul & Stoner, 2010; Kaya & Kafaligonul, 2019; Ogulmus et al., 2018). Observers were instructed to fixate during a trial and to indicate, by pressing one of two keys on a standard keyboard, which of the consecutive apparent motions appeared to move faster (i.e., two-interval forced-choice paradigm). Participants were allowed to respond at the end of each trial with no time pressure.

As in previous research (Kaya & Kafaligonul, 2019; Ogulmus et al., 2018), there was no additional task in the neutral (baseline) condition. The observers were asked to distribute their attention to all moving objects in the visual field and to make a comparison based on the overall speed (see also Table 1 for a comparison of attention conditions). The participants were informed that clicks would accompany the moving objects but to base their responses solely on the visual stimulation. In the cued condition, a brief (70 ms) square cue (0.5° length, blue: 20.4 cd/m^2 or red: 35 cd/m^2) was presented before the first apparent motion presentation (Fig. 1a). The cue location was at the center of one of the upcoming moving object's trajectory. The onset timing (i.e., onset asynchrony) between the cue and the first apparent motion was varied between 270 and 300 ms. The range of cue timing was selected to have sustained attention along the path of one of the moving objects (Nakayama & Mackeben, 1989; Ward, 2008). The observers were instructed to attend only to the moving object that would appear at the cue location and to compare the speed of that particular object. They also performed a secondary task by reporting the cue color. Since the cue was presented even before the first apparent motion, this secondary task was included in the design just to make sure that observers did not ignore the cue and they oriented attention at a specific location. In the fixation (color) condition, the observers were instructed to distribute their attention in the visual field and judge the overall speed as in the neutral condition. However, during the presentation of each apparent motion, the fixation color was turned to either red or green for 70 ms (see also Fig. 1a), and the onset of color change was varied within the visual time interval (ISI$_v$ = 100 ms). As a secondary task, the participants were also asked to report whether the fixation color change was the same or not. Since
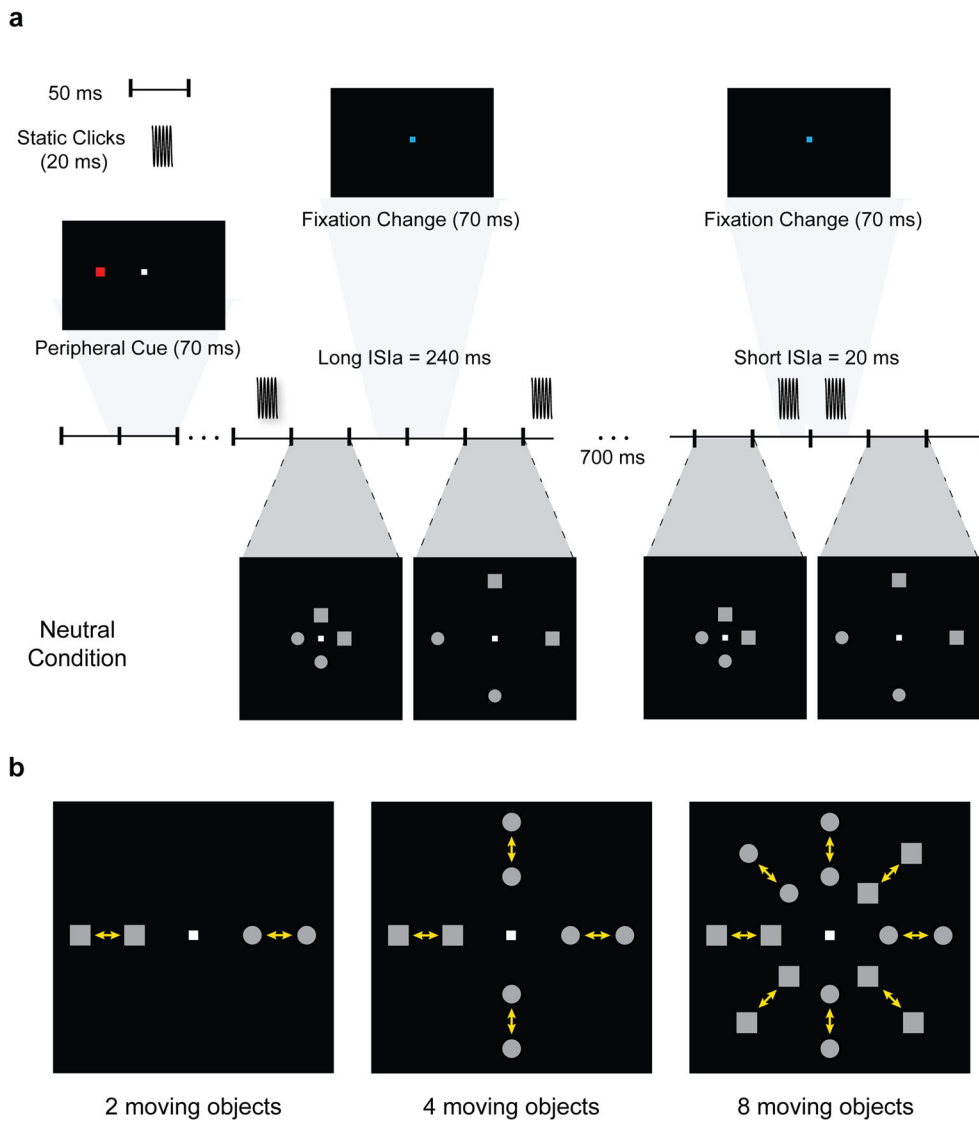
Fig. 1 a Representation of apparent motion frames and the timeline of stimulation for each attention condition. A two-frame apparent motion was presented twice during each trial with a temporal delay (ISI) of 700 ms. These consecutive visual apparent motions were precisely the same. However, the time interval between static clicks was either shorter (ISI$_a$ = 20 ms) or longer (ISI$_a$ = 240 ms) than the time interval between apparent motion frames (ISI$_v$ = 100 ms). Each motion frame had a 50 ms duration and included a specific number of objects (either 2, 4, or 8). Only the 4 moving object condition is displayed in the figure. Observers were asked to report which apparent motion (first or second) moved faster. In the neutral (baseline) attention condition, they were asked to distribute attention in the visual display. A peripheral cue was presented 270–300 ms before the onset of the first apparent motion for 70 ms in the cued condition. For the fixation condition, the fixation changed color for 70 ms during the presentation of each apparent motion. The changes in these two conditions were displayed above the timeline. The observers performed an additional secondary task on cue color/fixation color change in these conditions. **b** Spatial configurations of the different number of moving objects in the visual field. For each trial, all the moving objects were either in the outwards or inwards direction. The yellow arrows highlight the apparent motion paths, and they were not present on the actual display. (Color figure online)

the fixation color change occurred during the presentation of each apparent motion, the secondary task in this condition was included in the design to specifically manipulate attentional resources in the visual field and divert attention away from the moving objects. These three attention conditions (neutral, cued, and fixation) were run in separate blocks. The order of these blocks was randomized across participants. Each block consisted of 384 trials (3 different number of moving objects x 128 trials per condition).

Training and performance testing

Before the main behavioral experiment, each participant first engaged in practice/training blocks. These blocks allowed us to evaluate whether a participant can reliably compare the speed of two successive apparent motions in our experimental design and settings. There were no auditory clicks in the practice blocks, and the number of objects in each apparent motion frame was fixed to four (i.e., 4 moving object condition of the

Table 1 List and comparison of all attention conditions used in the study (the conditions of each experiment are grouped in separate rows)

Exp #		Attention conditions	Task
1	Neutral	2, 4, or 8 objects were presented in a two-frame apparent motion paradigm along with task-irrelevant sounds.	Speed Comparison
	Peripheral Cue	One of the objects was also cued prior to the onset of the first motion.	Speed Comparison + Report Cue Color
	Fixation	The color of the fixation also changed during each apparent motion (i.e., during the time interval between apparent motion frames).	Speed Comparison + Report Fixation Change
2	Neutral	The neutral condition was the same as the one in Experiment 1.	Speed Comparison
	Cue 1	One of the moving objects was presented in color during both apparent motions.	
	Cue 2	The colored object is also cued prior to the first apparent motion.	
3	Number	1, 3, or 5 objects among 12 objects moved, while the others remained stationary.	Speed Comparison
4	Neutral	The neutral condition was similar to the one in Experiment 1. However, speakers were positioned at different locations: adjacent, medium, and far. Auditory stimuli were also delivered through one of the speakers randomly on each trial.	Speed Comparison
	Sound Task	The observers performed an additional task on sound location.	Speed Comparison + Report Sound Location

main experiment; Fig. 1). As in previous research (Kafaligonul & Stoner, 2010; Ogulmus et al., 2018), one of the two successive apparent motions was used as a "reference" stimulus. The reference had a 100 ms time interval between apparent motion frames (ISI$_{ref}$ = 100 ms). The other "test" apparent motion had a time interval (ISI$_{test}$) that varied pseudorandomly from trial to trial: 20, 40, 60, 80, 100, 120, 140, 160, 180, and 200 ms. As in the main experiment (Fig. 1), the reference and test stimuli were separated by a delay of 700 ms, and their order was randomized from trial to trial. The reference and test apparent motions were not distinguished in the instructions to the participants. At the end of each trial, participants performed a speed comparison by indicating which apparent motion (i.e., first or second motion) appeared to move faster.

A practice block included a total of 120 trials (10 ISI$_{test}$ × 12 trials per condition). After each practice block, the percentage of trials in which the test apparent motion reported as faster was computed for each ISI$_{test}$ condition. The percentage of trials was expected to be high and above 75% for short ISIs (i.e., ISI$_{test}$ << ISI$_{ref}$). The percentage values should have decreased as the ISI$_{test}$ got longer and was expected to be below 25% for the long ISIs (i.e., ISI$_{test}$ >> ISI$_{ref}$). These percentage values were plotted as a function of ISI$_{test}$ and a complementary error function $(1-\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt)$ was fitted to these values using *psignifit* (Version 2.5.6). The software package implements the maximum likelihood method described by Wichmann and Hill (2001a, 2001b). The 50% point on the resultant curve yields the point of subjective equality (PSE). The PSE is the ISI$_{test}$ for which the test apparent motion was reported as faster than the reference on 50% of the trials (see also Fig. S1 for sample data). To be eligible to continue with the main experimental session, we required that the PSE point was reliably estimated based on the data for the whole ISI$_{test}$

range (20–200 ms). We expected the percentage values of two short ISI$_{test}$ conditions (slower test: 20, 40 ms) to be above or equal to 75% and two long ISI$_{test}$ conditions (faster test: 180, 200 ms) to be below or equal to 25% level. We also required the values in three of these four extreme ISI$_{test}$ conditions to be in the expected range. Participants were trained by repeating the practice block until they reached these criteria.

Results

The results of Experiment 1 are shown in Fig. 2. To quantify auditory time interval effects on perceived speed, we computed the percentage of trials in which the apparent motion with a short auditory time interval was perceived to move faster than the one with a long auditory interval. In all the experimental conditions, the mean percentage values were above the 50% chance level (Fig. 2a). A series of one-sided one-sample permutation tests (sampling permutation distribution 5k) were performed on the percentage value of each condition to assess whether these values were greater than the chance level. The resultant p values were corrected with the Holm method for nine comparisons (i.e., 3 attention conditions × 3 number of objects). All the data analyses were performed in R (Version 4.1.2; R Core Team, 2021). The results showed that for all the conditions the percentage values were significantly higher than 50% (neutral: $p_{adj} < .001$, $p_{adj} = .0024$, $p_{adj} = .0016$; cue color: $p_{adj} = .0016$, $p_{adj} = .0032$, $p_{adj} = .0054$; fixation color: $p_{adj} = .003$, $p_{adj} = .0032$, $p_{adj} = .0032$ for 2, 4, and 8 objects, respectively). These results indicate reliable temporal ventriloquism (i.e., auditory time interval) effects on perceived visual speed in all the conditions tested.

According to a Shapiro–Wilk test, residuals of percentage values of apparent motion perceived as faster were not normally distributed ($W = 0.95$, $p < .001$). Additionally, for
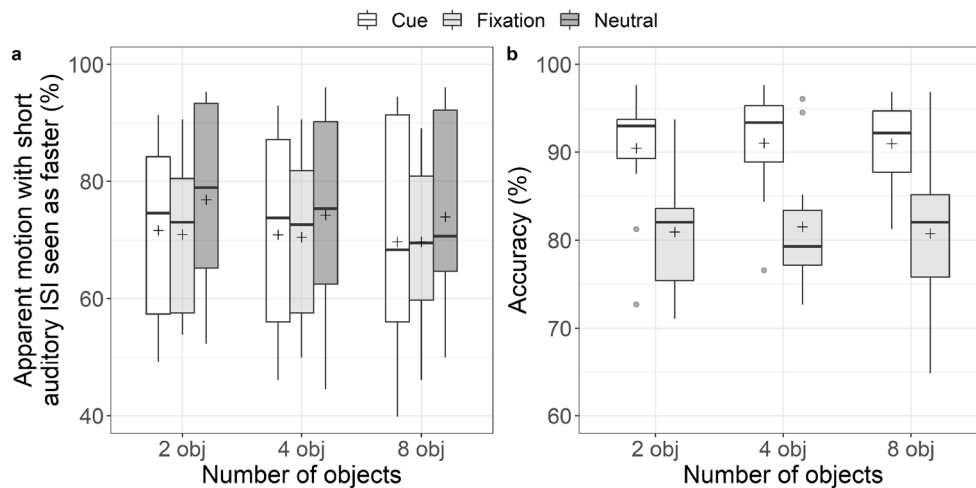
Fig. 2 Results of Experiment 1 (*n* = 12). **a** Boxplots of the percentage of trials in which the apparent motion with a short auditory time interval reported as faster is displayed for each condition. Each attention condition is represented by a distinct gray level, and the boxplots of each number of moving objects are grouped together. **b** Boxplots of performance values for the secondary task. For each boxplot, the horizontal black line indicates the median, and the lower and upper hinges correspond to the first and third quartiles (i.e., the 25th and 75th percentiles). Instead, the plus sign within each boxplot represents the mean percentage (**a**) and mean accuracy (**b**) values. The upper whisker extends from the hinge to the largest value no further than 1.5 * IQR of the hinge (where IQR is the interquartile range or distance between the first and third quartiles). The lower whisker extends from the hinge to the smallest value at most 1.5 * IQR of the hinge. The gray points in panel (**b**) indicate outliers

Experiment 1, data are likely to follow a uniform distribution (data distribution was assessed using the R function *descdist* with 1 k bootstrapped values). Therefore, we used the aligned rank transform (ART), a procedure for the nonparametric analysis of variance in multifactor designs (Higgins et al., 1990; Higgins & Tashtoush, 1994; Salter & Fawcett, 1993; Wobbrock et al., 2011). With this technique, a linear mixed model can be implemented once the data is aligned and ranked for each main and interaction effect. Pairwise comparisons were conducted using the ART-C procedure (Elkin et al., 2021). A linear mixed model with random intercept across participants and including the attention conditions (neutral, cue color, and fixation color) and the number of objects (2, 4, and 8) as within-subjects factors, revealed only a significant effect of attention conditions, $F(2, 88) = 4.55$, $p = .013$, number of objects: $F(2, 88) = 0.27$, $p = .77$; interaction between attention and number of objects: $F(2, 88) = 0.089$, $p = .98$. For the main effect of the attention, Holm-corrected post hoc comparisons reported a significant difference between the neutral and the cue color condition ($p_{adj} = .028$), between the neutral and the fixation color condition ($p_{adj} = .024$), but not between cue color and fixation color condition ($p_{adj} = .84$).

Figure 2b shows the averaged performance values for the secondary task. Participants reported either the cue color or had to discriminate the color change of the fixation square. A series of one-sided one-sample permutation tests (sampling permutation distribution 5k) were performed for each condition on the accuracy values to assess whether accuracies across conditions and number of objects were greater than 75%. The results showed that for all the conditions the percentage values were significantly higher than 75%

(Holm-corrected comparisons; cue color: $p_{adj} = .0012$, $p_{adj} = .002$, $p_{adj} = .002$; fixation color: $p_{adj} = .0116$, $p_{adj} = .0036$, $p_{adj} = .021$, for 2, 4, and 8 objects, respectively). For accuracy values, residuals were not normally distributed ($W = 0.94$, $p = .0024$). Therefore, we again used the ART with a linear mixed model. The analysis revealed only a significant effect of the attention condition, $F(2, 55) = 80$, $p < .001$; number of objects: $F(2, 55) = 0.047$, $p = .95$; interaction between attention condition and number of objects: $F(2, 55) = 0.16$, $p = .85$. Overall, the accuracy values suggest that observers attended to the cue location or fixation target and performed the secondary task according to the instructions.

Discussion

The auditory time interval effects on perceived speed were mainly present in all conditions, and the results did not indicate a significant effect of the number of moving objects. Given that the effects of auditory time intervals have been mostly described as each click altering the perceived timing of each apparent motion frame, these findings suggest that the timing of a single auditory click can drive the timing of more than one object presented in the visual field. There was a significant main effect of attention. However, compared with the neutral condition, the auditory time interval effects were significantly lower when observers attended to a single moving object in the visual field. Based on the hypothesis that there is a limited capacity for the number of visual events that can be bound to a single auditory event, we expected to have

higher percentage values (Fig. 2a) for the cued condition in which observers attended to a single moving object. More importantly, these results revealed a significant effect of perceptual load/attention demands in the visual field. In the fixation condition, we diverted attention to a stationary object (i.e., fixation target) during the presentation of each apparent motion. According to the previous research (Alsius et al., 2005; Ren et al., 2020; Ren et al., 2021), we expected a decrease in the amount of audiovisual interactions and hence to have lower percentage values in this condition compared with the neutral condition. In line with this original prediction, the percentage values for the fixation condition were significantly lower than those of the neutral condition.

Experiment 2

Against the original prediction, a spatial cue did not improve auditory time interval effects on perceived visual speed in the previous experiment. The spatial attention was manipulated in a goal-driven manner (Theeuwes & Failing, 2020) by using a static cue and introducing a secondary task relevant to the cue. Although the participants were instructed carefully, it is still conceivable that they might have allocated their attention to the cue itself rather than to the moving object at the cued location. Moreover, high perceptual load due to the discrimination and then speed comparison in a dual-task paradigm might have overshadowed any potential cueing effects in the spatial domain. For instance, having a secondary task on cue color (i.e., an object other than the moving stimuli) might have decreased audiovisual interactions. This decrease might have canceled out any enhancement due to cueing and allocation of attention at the specific location of the moving object. Hence, the spatial cue together with a secondary task, might not efficiently modulate temporal ventriloquism effects on perceived speed. To address these concerns and restrict the contribution of other confounding factors, we re-examined a potential modulatory role of spatial cueing by using a simplified experimental procedure and without having a secondary task in a control experiment.

Methods

Participants

Ten naïve volunteers (age range: 21–23 years) participated and completed all experimental procedures.

Stimuli and procedure

The apparent motion stimulation, number of visual objects, auditory clicks, and timeline of events during a trial were the same as those in Experiment 1. There were three primary attention conditions that were run in separate blocks (Table 1). As in Experiment 1, participants were instructed to distribute their attention to all moving objects in the visual field and to make a comparison based on the overall speed in the neutral (baseline) block. In the second condition (i.e., Cue 1 condition), we manipulated attention in a stimulus-driven manner by displaying one of the moving objects in red. The observers were instructed to attend to the red object and compare the speed of that object. In the third condition (i.e., Cue 2 condition), there was an additional red cue (0.55° length square, 35 cd/m²) prior to the apparent motion frames, which informed about the location of the red object in the visual display. Similar to the previous experiment, the cue duration was 70 ms, and it appeared 300 ms before the first apparent motion (onset-to-onset timing). This third attention condition included both the visuospatial cue from Experiment 1 and the stimulus-driven component implemented by presenting one object in a different color to make it distinct among the other objects. Accordingly, the overall cueing effect was expected to be stronger in this condition. There was no additional/secondary task in any conditions of the experiment, and the observers only compared the speed of consecutive apparent motions and reported which one was faster.

In Experiment 1, against instructions, observers could have conceivably ignored apparent motions and relied only on auditory time intervals for the speed comparison. Although this is unlikely due to the procedure used in training/practice blocks (see Experiment 1: Training and performance testing), catch trials were also included in this experiment to ensure that observers performed speed judgement according to the instructions. In the catch trials, the auditory time intervals of two consecutive presentations were the same ($ISI_a = 100$ ms). However, the visual time intervals ($ISI_v = 20$ ms or 180 ms) were different to have fast and slow apparent motions during a trial. These time intervals were adjusted to have a reliable difference between the speed of two apparent motions even in the presence of auditory clicks with a 100 ms interval. The order of fast ($ISI_v = 20$ ms) and slow ($ISI_v = 180$ ms) apparent motions was randomized across trials. An observer who performed the perceptual task according to the instructions was expected to typically report the apparent motion with 20 ms ISI_v as faster than the one with 180 ms ISI_v. On the other hand, an observer who just relied on auditory click timing rather than visual speed should not have reported a difference between apparent motions and hence, had a performance value around the chance level (i.e., 50% level in the two-interval forced-choice paradigm). A total of 96 catch trials were used in an experimental session. These trials were mixed with the main trials, and they were not distinguished in the instructions to the observers. All other stimulus parameters, experimental conditions, and procedures (including practice blocks and performance criteria) were the same as those in Experiment 1.
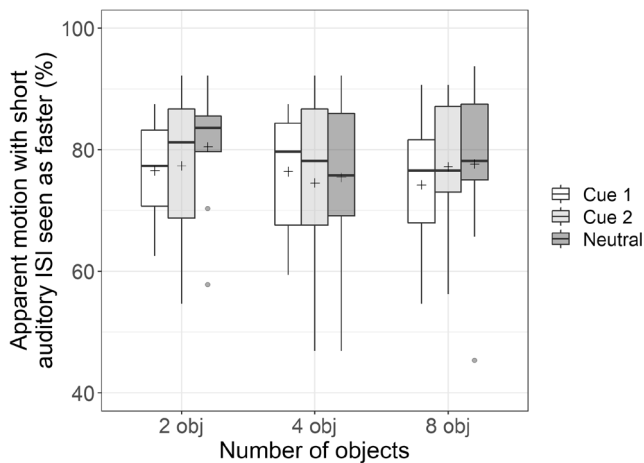
Fig. 3 Results of Experiment 2 (*n* = 10). Boxplots of the percentage of trials in which the apparent motion with a short auditory time interval reported as faster is displayed for each condition. Each attention condition is represented by a distinct gray level, and the boxplots of each number of moving objects are grouped together. For each boxplot, the horizontal black line indicates the median, and the lower and upper hinges correspond to the first and third quartiles (i.e., the 25th and 75th percentiles). The gray points indicate outliers. Other conventions are the same as those in Fig. 2

Results

The percentage of trials in which the apparent motion with a short interval seen as faster is shown in Fig. 3. As in Experiment 1, a series of one-sided one-sample permutation tests (sampling permutation distribution 5 k) were applied to the percentage value of each condition to assess whether each percentage value was greater than the chance level (50%). The results showed that for all the conditions the percentage values were significantly higher than 50% (Holm-corrected comparisons; neutral: p_{adj} = .0088, p_{adj} = .0088, p_{adj} = .0088; Cue 1: p_{adj} = .0054, p_{adj} = .007, p_{adj} = .008; Cue 2: p_{adj} = .007, p_{adj} = .0088, p_{adj} = .0064 for 2, 4, and 8 objects, respectively). These results indicate significant effects of auditory time intervals on perceived visual speed in all the conditions tested in Experiment 2.

A Shapiro–Wilk test showed that residuals for percentage values were not normally distributed (*W* = 0.93, *p* < .001), with a negative skewness of −0.8 (*SE* = 0.25). Using the median absolute deviation with a cutoff of 3 (Leys et al., 2013), we also identified four outliers that were included in the analysis (percentage values <50%). Data were analyzed using a generalized linear model (GLM; Fox, 2003) with lme4 package (Bates et al., 2015). A Gamma function and identity link transformation function were used in the GLM model. We chose a Gamma function for the regression analysis because almost all the percentage values fell into the Gamma quantiles, allowing to deal with outliers without removing them or

transforming the original data (Zuur et al., 2010) and because data distribution was well approximated by a Gamma distribution. The identity link transformation function means that percentage values were not transformed. The model included the attention conditions (i.e., neutral, Cue 1, and Cue 2), the number of moving objects, and the interaction between attention and the number of moving objects as predictors. The regression analysis did not report any significant main effect or interaction (attention: χ^2 = 0.442, *df* = 2, *p* = .802; number of moving objects: χ^2 = 0.681, *df* = 2, *p* = .711; attention × number of moving objects: χ^2 = 0.651, *df* = 4, *p* = .957). The coefficients of the regression analysis are reported in Table S1 (Supplementary Material).

In the catch trials, the auditory time interval was fixed at 100 ms, but the time interval between the apparent-motion frames (ISI_v) differed. For each condition (i.e., 3 attention conditions × 3 number of objects), we computed the percentage of trials in which the apparent motion with a short visual interval was perceived as faster. As expected, the mean percentage values were much higher than the 50% chance level (see Fig. S2 in the Supplementary Material). A series of one-sided one-sample permutation tests (sampling permutation distribution 5 k) were performed on the percentage value of each condition to assess whether these percentages were significantly higher than 65%. The results showed that for all the conditions the percentage values were significantly higher than 65% (Holm-corrected comparisons; neutral: p_{adj} = .0108, p_{adj} = .0224, p_{adj} = .0072; Cue 1: p_{adj} = .0224, p_{adj} = .0108, p_{adj} = .0224; Cue 2: p_{adj} = .0072, p_{adj} = .0098, p_{adj} = .0224 for 2, 4, and 8 objects, respectively). According to a Shapiro–Wilk test, residuals of these percentage values were not normally distributed (*W* = 0.914, *p* < .0001). Additionally, the data were likely to be uniformly distributed. Again, we used the Aligned Rank Transform (ART). A linear mixed model with random intercept across participants and including the attention condition (neutral, Cue 1, and Cue 2) and the number of objects (2, 4, and 8) as within-subjects factors, did not reveal any significant main effect or interaction, attention condition: *F*(2, 72) = 0.18, *p* = .83; number of objects: *F*(2, 72) = 1.19, *p* = .31; interaction between attention and number of objects: *F*(2, 72) = 1.27, *p* = .29. Overall, these high percentage values confirm that participants performed speed comparison according to the instructions and rule out any decisional bias on auditory time intervals, such as only relying on auditory time intervals and ignoring visual motions while performing the task.

Discussion

Compared with the neutral (i.e., distributed attention in the visual field) condition, we expected an enhancement in

audiovisual binding and thus in interactions when attention was allocated to a moving object at a specific location. Therefore, the cued conditions were expected to have larger percentage values. In contrast to this prediction, the percentage values were around the same level across conditions. Moreover, in all the conditions, the temporal ventriloquism effects on perceived speed were present. These findings confirm the existence of audiovisual interactions regardless of the number of moving objects and highlight the automatic nature of these interactions.

Experiment 3

In the previous experiments, we investigated the relationship between the number of moving objects and the amount of audiovisual interactions by systematically manipulating the number of concurrent objects in apparent motion frames. The random assignment of shapes (circles and squares) to the locations with different angles on imaginary circles led to a final percept of moving objects in different directions. This was particularly achieved when there were two moving objects in the visual field. In this condition, two distant objects with different shapes moved in the opposite directions (Fig. 1b). The possibility of any grouping and inducing a global motion percept was low, and the design led to a percept of more than one moving object in the visual field. The neutral condition of two moving objects provided a baseline/test condition not only for testing the basic hypothesis that audiovisual binding is limited to one moving object but also for understanding the effects of spatial cueing/attentional demands. By including 4 and 8 moving objects in the design, we wanted to further characterize the dependency of temporal ventriloquism on the number of moving objects in the visual field. On the other hand, for the 4- and 8-object moving conditions, it is still possible that an orderly presentation of objects in the cardinal and diagonal directions may engage the grouping of objects in the spatial domain. That is, the participants might have experienced single and integrated motion in the visual field. Thus, the timing of a single click may influence the perceived speed even if the number of physical objects increases in each motion frame. To rule out this possibility, we designed an additional control experiment based on the original paradigm by Van der Burg et al. (2013). We used 12 objects in the visual field, and only a portion of them moved (randomly selected 1, 3, or 5 objects). The remaining objects were static and acted as background. The static ones efficiently broke down any integration in the whole visual field and led to a percept of distinct moving objects in different directions.

Methods

Participants

Nine naive volunteers (age range: 19–30 years) participated and completed all procedures of the experiment. One of the observers took part in Experiment 2.

Stimuli and procedure

We used the basic stimulus parameters, conditions, and procedures of Experiment 2. However, 12 objects (circles or squares) were equally spaced around the fixation target on an imaginary circle with a radius of 4.7°. Based on the number of moving objects (1, 3, or 5), some of these locations were selected randomly. The selected ones were 3.85° and 5.55° away from the fixation point (rather than 4.7°) in each apparent motion frame. In other words, the selected ones were used to generate moving objects, and the remaining ones were static and positioned in the middle of the apparent motion path at a different angle on the imaginary circle (Fig. 4; see also Table 1). The motion direction was selected randomly for each trial, and all the moving objects were either in the outwards or inwards direction.

Only the neutral (baseline) attention condition of Experiment 2 was used. The participants were instructed to distribute their attention in the visual field and asked to compare the overall speed of two successive presentations. There was no secondary task. Each participant completed a session of 384 trials (3 different number of objects × 128 trials per condition) and 96 catch trials. All other experimental procedures, practice/training blocks, and inclusion/exclusion criteria were the same as those in Experiment 1.

Results

The percentage of trials in which the apparent motion with a short auditory interval perceived as faster is shown in Fig. 5. As in Experiments 1 and 2, a series of one-sided one-sample permutation tests (sampling permutation distribution 5k) were performed to assess whether each percentage value was significantly higher than the chance level (50%). The results showed that the percentage values of all conditions were significantly higher than 50% (Holm-corrected comparisons, all $p_{adj} = .0054$).

A Shapiro–Wilk test showed that residuals for percentage values of apparent motion with the short auditory interval perceived as faster were normally distributed ($W = 0.967$, $p > .05$). Two outlier data points were identified (percentage values >60%) and included in the analysis. A repeated-measures ANOVA did not reveal a significant effect of the number of moving objects, $F(1.24, 13.23) = 0.276$, $p = .661$,
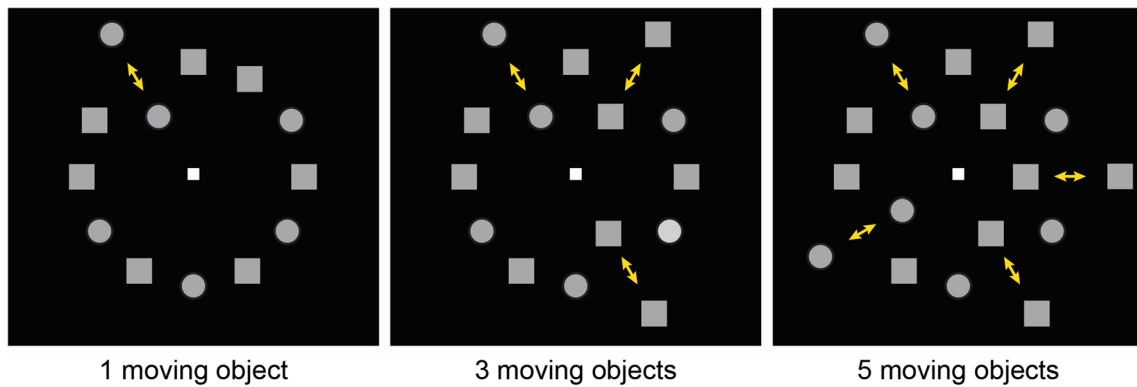
Fig. 4 Spatial configurations for different moving object conditions in Experiment 3. Based on the number of moving objects (1, 3, or 5), some of the locations/angles on the imaginary circle were selected randomly. The selected locations were used for moving objects, and the static objects were positioned at the remaining ones. For each trial, all the moving objects were either in the outwards or inwards direction. The yellow arrows highlight the apparent motion paths, and they were not present on the actual display. (Color figure online)

$\eta_p^2 = 0.033$. Given that the sphericity assumption was violated ($p = .038$) degrees of freedom were corrected using the Greenhouse–Geisser correction.

In catch trials, the observers typically reported the apparent motion with a short visual time interval as faster (see Fig. S3 in the Supplementary Material). A series of one-sided one-sample permutation tests (sampling permutation distribution 5k) were performed for each number of objects to assess whether the percentage values were significantly higher than 65%. Holm-corrected comparisons showed that for 3 and 5 moving objects, the percentage values were significantly higher than 65% ($p_{adj} = .0048$ and $p_{adj} = .0096$ for 3 and 5 moving objects, respectively), but not for one moving object ($p_{adj} = .115$). However, the percentage value of one moving object was significantly higher than the 50% chance level ($p_{adj} = .0054$). According to a Shapiro–Wilk test, residuals of

percentage values of apparent motion perceived as faster were normally distributed ($W = 0.937$, $p > .05$). We found that the number of moving objects significantly affected these percentage values, $F(2, 16) = 9.35$, $p = .002$, $\eta_p^2 = 0.54$. The percentage value for the one moving object condition was significantly lower than those of the conditions with 3 and 5 moving objects (Holm-corrected post hoc comparisons, all $p_{adj} < .05$).

Discussion

In this experiment, we wanted to re-examine whether the timing of a brief static click can drive the timing of more than one moving object in each motion frame, and hence the auditory time interval affect the speed perception of more than one moving object. The results indicated reliable and robust auditory time interval effects over multiple and simultaneous moving objects. Moreover, there was no significant main effect of number of moving objects on these audiovisual interactions in the temporal domain. Interestingly, we found a significant effect of number of moving objects in the catch trials. Although these trials were designed to ascertain any basic decisional bias on auditory time intervals, they do not preclude temporal ventriloquism since there was a mismatch between auditory and visual time intervals. The decrease in the percentage value of one moving object condition might indicate an increase in the effects of auditory time intervals on the final percept. Accordingly, this decrease might suggest an enhancement of audiovisual interactions and facilitation of binding when the number of visual objects is one. However, this possibility was not supported by the catch trials of other experiments and the main trials of the current experiment.
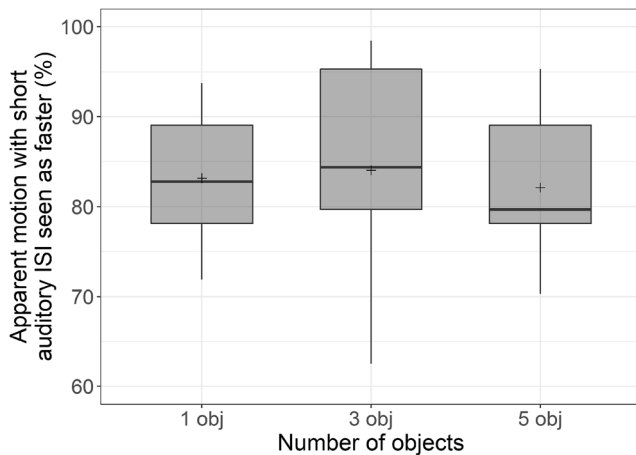
It is also important to note that the location on the imaginary circle and shape of all objects were randomly assigned from trial to trial. When there were 3 and 5 moving objects, the randomization and the presence of static objects efficiently broke down any global motion percept. The selected objects

Fig. 5 Results of Experiment 3 ($n = 9$). Boxplots of the percentage of trials in which the apparent motion with a short auditory time interval reported as faster for each number of moving object condition. For each boxplot, the horizontal black line indicates the median, and the lower and upper hinges correspond to the first and third quartiles (i.e., the 25th and 75th percentiles). The plus sign within each boxplot represents the mean percentage value

with random shapes were distinctively moved in different directions and led to an efficient neutral/distributed attentional condition. On the other hand, the first frame of a single moving object in the visual field might be distinguished and conceivably capture attention to a single location even if its location was randomized. Against our instructions, the observers might have involuntarily allocated attention to a particular location in the visual field. Even this case would provide an important control condition to test the hypothesis that audio-visual binding is limited to one moving object. In this specific condition, temporal ventriloquism effects on perceived speed were expected to be higher. However, compared with other conditions, there was no improvement and the observed effects were around the same level. Overall, our findings did not provide any convincing evidence for the hypothesis that there is a limited capacity for the number of visual events that can be bound to a single auditory event. They rather suggest efficient processing and binding in complex audiovisual stimulations (see also Wilbiks & Dyson, 2016, 2018)

Experiment 4

In the previous experiments, we investigated the effects of spatial attention and attentional demands in the visual field. The findings revealed a significant role of attentional demands/perceptual load. To complement these findings in the auditory domain, we examined whether the allocation of attentional demands in the auditory space has a role in the observed effects of temporal ventriloquism. While interpreting the effects of auditory time intervals on motion perception, the audition has been considered as the dominant modality (i.e., capturing modality) in the temporal domain (Chen & Vroomen, 2013). Therefore, we hypothesized that allocating attention to this dominant modality would facilitate auditory signals and associated processes, and hence increase the observed auditory time interval effects on perceived speed. To test this hypothesis, we used a similar dual-task paradigm, but the secondary task was based on the spatial position of static clicks rather than an object in the visual field. In addition, we manipulated the secondary task difficulty in the auditory space by having distinct conditions of click position.

Methods

Participants

Nine naive volunteers (age range: 19–29 years) participated and completed all procedures of the experiment. Two of these observers took part in Experiment 1, and one of the observers participated in both Experiments 2 and 3.

Stimuli and procedure

The visual stimulation, auditory clicks, experimental design, and timeline of events during a trial were the same as those described in Experiment 1. Rather than a binaural presentation, the clicks were presented either from the right or left speaker. The location (left vs. right) was randomized across trials. The distance between the speakers was pseudorandomly selected from three values (center-to-center horizontal distance, adjacent: 8 cm, middle: 35 cm, far: 62 cm) and was fixed during an experimental block. Each block consisted of 240 trials (3 different number of objects × 80 trials per condition) and 48 catch trials (3 different number of objects × 16 trials per condition).

In the neutral (baseline attention) condition, observers were asked to fixate during a trial and to perform a speed comparison task (i.e., to indicate whether the first or second apparent motion appeared faster) at the end of a trial. In the auditory attention condition, there was an additional secondary task in which participants reported the location of clicks (left vs. right) by pressing one of the keys on a standard keyboard (Table 1). We also manipulated the secondary task difficulty by having a systematic change in the distance between the speakers. The attention and speaker location conditions (2 attention conditions × 3 speaker locations) were run in 6 separate blocks. Data were collected within the same day by randomizing the order of blocks across participants.

Results

The percentage values of the main trials are shown in Fig. 6. As in the previous experiments, a series of one-sided one-sample permutation tests (sampling permutation distribution 5k) were performed to assess whether the percentage values of apparent motion perceived as faster were significantly higher than the chance level (50%). Permutation tests were performed separately for each speaker position so that the resultant p values were Holm-corrected for six comparisons (i.e., 2 conditions [neutral vs. attention to sound] × 3 number of moving objects). The results showed that the percentage values were significantly higher than 50% across all the conditions (see Table S2 in the Supplementary Material). These results indicate reliable effects of auditory time intervals on perceived visual speed in all the conditions tested.

A Shapiro–Wilk test showed that residuals for percentage values of the apparent motion with a short auditory interval perceived as faster were not normally distributed ($W = 0.97$, $p < .01$), with a negative skewness of −0.1 ($SE = 0.19$). Using the median absolute deviation with a cutoff of three (Leys et al., 2013), we identified one outlier that was included in the analysis (percentage value <50%). Additionally, data were likely to follow a uniform distribution. Therefore, we used the
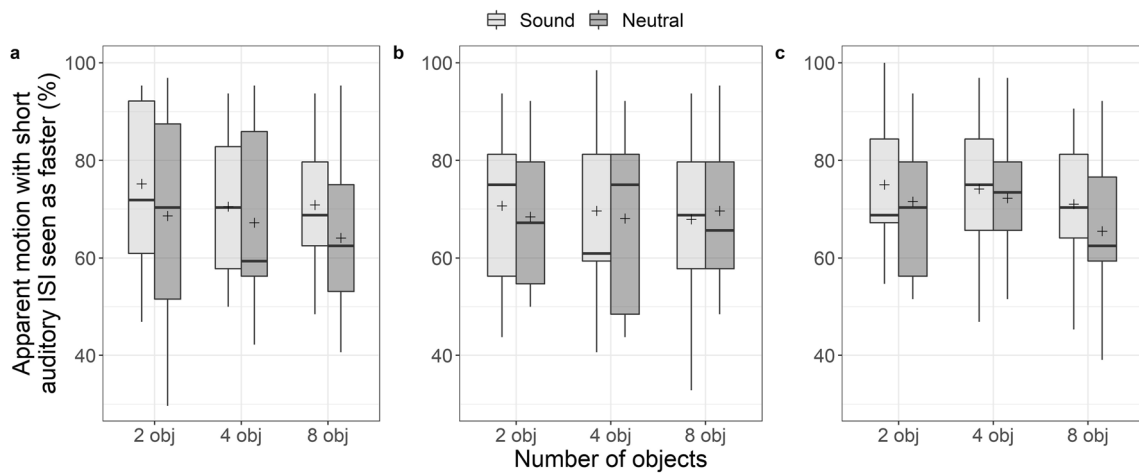
Fig. 6 Results of Experiment 4 (*n* = 9). Boxplots of the percentage of trials in which the apparent motion with a short auditory time interval reported as faster is displayed for each condition. Panel (**a**) indicates adjacent speaker position, panel (**b**) middle position, and panel (**c**) far position. In each panel, each attention condition (neutral vs. sound) is represented by distinct gray levels, and the boxplots of each number of moving objects are grouped together. For each boxplot, the horizontal black line indicates the median, and the lower and upper hinges correspond to the first and third quartiles (i.e., the 25th and 75th percentiles). The plus sign within each boxplot represents the mean percentage

ART. A linear mixed model with random intercept across participants and including the speaker position (adjacent, middle, and far), attention condition (neutral and attention to sound), and the number of moving objects (2, 4, and 8) as within-subjects factors, revealed only a significant effect of the attention condition, $F(1, 136) = 7.65$, $p = .006$. All other main effects and interactions were not significant ($p > .05$). These results suggest that when participants had to allocate attention to the sound, the percentage values, and hence temporal ventriloquism effects on speed perception increased.

The accuracy values for locating the auditory clicks (left vs. right speaker) are shown in Fig. 7. A series of one-sided one-sample permutation tests (sampling permutation distribution 5k) were performed to assess whether the accuracy values in the secondary task were significantly higher than the 75%. Permutation tests were performed separately for each speaker position (adjacent, middle, and far), so that the resultant *p*-values were Holm-corrected for three comparisons (i.e., 3 number of moving objects per speaker position). The results showed that the accuracy values were significantly higher than
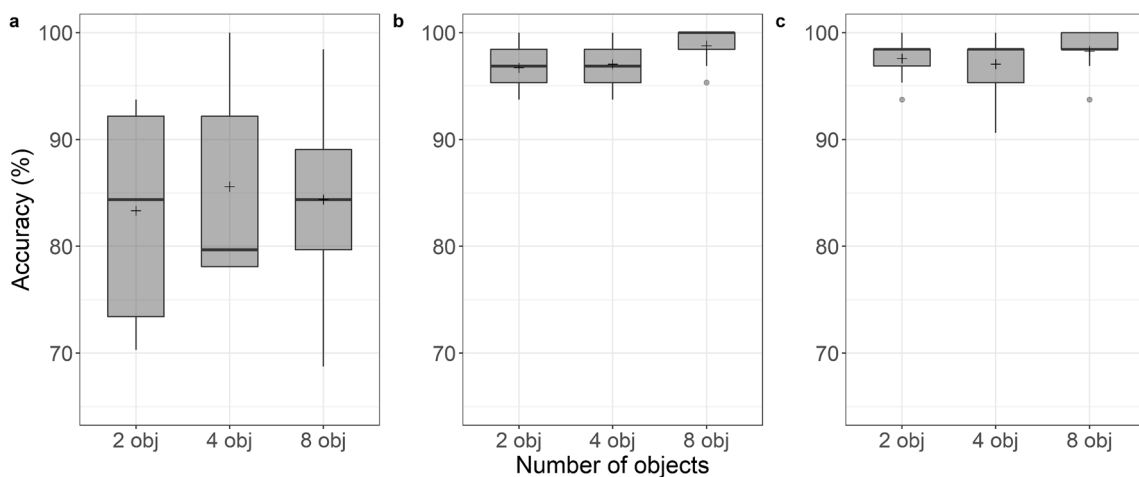

Fig. 7 Boxplots of the accuracy values for the secondary task on click location in Experiment 4 (*n* = 9). Panel (**a**) indicates adjacent speaker position, panel (**b**) middle position, and panel (**c**) far position. For each boxplot, the horizontal black line indicates the median, and the lower and upper hinges correspond to the first and third quartiles (i.e., the 25th and 75th percentiles). The plus sign within each boxplot represents the mean accuracy. The gray points in panels (**b**) and (**c**) indicate outliers

75% across all the speaker positions (adjacent: all $p_{adj} < .05$; middle: all $p_{adj} < .01$; far: all $p_{adj} = .003$). A Shapiro–Wilk test showed that residuals were not normally distributed ($W = 0.769$, $p < .001$), with a strong negative skewness of -1.445 ($SE = 0.267$). Using the median absolute deviation with a cutoff of three (Leys et al., 2013), we identified 13 outliers that were included in the analysis (percentage value >50%). Therefore, we used ART procedure with a linear mixed model including random intercept across participants and speaker position (adjacent, middle, and far) and number of moving objects (2, 4, and 8) as within-subjects factors. The analysis revealed a significant effect of the speaker position, $F(2, 64) = 54.38$, $p < .001$, but not a significant effect of the number of moving objects, $F(2, 64) = 1.81$, $p = .17$, or an interaction between speaker position and number of objects, $F(4, 64) = 0.33$, $p = .86$. Holm-corrected post hoc comparisons for the speaker position reported a significant difference between adjacent and middle speaker positions ($p_{adj} < .001$) and between adjacent and far speaker positions ($p_{adj} < .001$), but not between middle and far speaker positions ($p_{adj} = .63$).

In catch trials, observers typically reported the apparent motion with a short visual time interval as faster (see Fig. S4 in the Supplementary Material). Permutation tests (sampling permutation distribution 5 k) were performed separately for each speaker position, so that the resultant p values were Holm-corrected for six comparisons (i.e., 2 conditions [neutral vs. attention to sound] × 3 number of moving objects). The results showed that these percentage values were significantly higher than 50% across all the conditions (see Table S3 in the Supplementary Material). A Shapiro–Wilk test showed that residuals for the percentage values of catch trials were not normally distributed ($W = 0.924$, $p < .0001$), with a negative skewness of -0.904 ($SE = 0.191$). Using the median absolute deviation with a cutoff of three (Leys et al., 2013), we identified seven outliers that were included in the analysis (five outliers >50% and two outliers <50%). The ART procedure with a linear mixed model did not reveal any significant main effect or interaction (all ps > .05).

Discussion

These findings complement the results of previous experiments on the visual field by revealing an effect of attentional demands/perceptual load in the auditory space. However, these modulations were in the opposite direction and facilitated the auditory time interval effects on perceived speed. When participants allocated attention to the clicks via a secondary task, the percentage values and thus temporal ventriloquism effects on speed perception increased. Accordingly, these modulations in the percentage values are in line with the original hypothesis. These results provide important evidence that allocation of attentional resources to the dominant modality (i.e., audition) in the temporal domain can facilitate

audiovisual interactions and their influences on speed perception. As in previous experiments, the outcome of catch trials confirmed that participants performed speed comparisons according to the instructions. The behavioral results also revealed a significant effect of speaker position on the accuracy scores of the secondary task, showing that task difficulty was successfully manipulated. However, neither the speaker position nor the elicited task difficulty was represented in the modulations of the percentage values of speed comparison.

General discussion

In four different experiments, we investigated the modulatory role of attention in audiovisual interactions in time. Accordingly, we used a design based on temporal ventriloquism (i.e., auditory time interval) effects on perceived speed. We oriented attention either in the visual or auditory domain and also changed the number of moving objects systematically. We did not find a significant and meaningful effect of spatial cueing in the visual field. On the other hand, introducing an additional task in the visual or auditory domain significantly modulated the amount of temporal ventriloquism effects on perceived speed. Therefore, these results revealed an important modulatory role of attention demands. Moreover, the effects of auditory time intervals on perceived speed were mostly constant across different number of moving objects and existed in all the experimental conditions. Thus, our findings also indicated that the time interval demarcated by static clicks can drive the perceived timing and speed of more than one moving object in the visual field.

Spatial cueing

Daily life situations mostly require the selection and prioritization of relevant information arising from different locations in the visual field. This also applies to visual motion processing. The selection process has particular importance to have correct estimates of direction and speed when there is more than one moving object in the visual field. An important question concerns whether orienting attention in the spatial domain modulates auditory time interval effects on perceived speed. In Experiment 1, the amount of these crossmodal effects on perceived speed significantly decreased when attention was oriented to a moving object at a specific location. However, based on the hypothesis that audiovisual binding is limited to a single visual event (Van der Burg et al., 2013), we particularly expected an enhancement of audiovisual interactions and hence an increase in the amount of temporal ventriloquism effects on perceived speed when observers focused on a single moving object. The results did not provide any supporting evidence for such an enhancement. In Experiment 2, we tested the effect of cueing by using more than one cue type and

without having a secondary task. The results did not indicate any significant effect of cueing in the visual field. When the outcome of both experiments is taken into consideration, we did not find a significant and meaningful effect of spatial cueing in the visual field. Overall, our results are in line with the initial findings on spatial ventriloquism. Consistent with the fact that vision has better spatial resolution than audition, a visual stimulus (e.g., flash) can attract and bias the perceived location of a primary sound (e.g., static click/tone) in this illusion. This analogous phenomenon provides an important demonstration of visual dominance in the spatial domain. Using paradigms based on spatial ventriloquism, several studies have shown that the amount of position shift (i.e., the attraction of perceived sound location toward the physical location of visual stimulus) is immune to the manipulations of endogenous and exogenous attention in the visual field (e.g., Bertelson et al., 2000; Vroomen et al., 2001a, 2001b). These audiovisual interactions in the spatial domain were present regardless of the focus of visual spatial attention, suggesting the automatic and stimulus-driven nature of crossmodal interactions. Our results here complement the previous findings on spatial ventriloquism by highlighting a similar nature of audiovisual interactions in the temporal domain.

Of particular relevance to the current study, the role of spatial attention in audiovisual interactions has been investigated with dynamic paradigms, including motion. Using a variant of the crossmodal dynamic capture paradigm (Soto-Faraco et al., 2002), Sanabria et al. (2007) quantified audiovisual interactions in motion and assessed the role of spatial attention in these interactions. In a typical crossmodal dynamic capture paradigm, the participants report the direction of an auditory apparent motion (primary modality) during the concurrent presentation of a visual apparent motion (secondary modality). As in spatial ventriloquism, the visual stimulation typically dominates in the spatial domain and thus biases the perceived direction of auditory motion. The direction discrimination performance for auditory motion significantly drops when the visual motion is presented in the opposite direction, compared with the condition in which auditory and visual apparent motions had the same direction. The dynamic capture effect is quantified by taking the performance difference between the two (same vs. opposite direction of visual motion) conditions. Sanabria et al. (2007) combined this design with endogenous and exogenous spatial cueing. The crossmodal dynamic capture effect was decreased in the cued trials, suggesting that spatial attention modulates audiovisual interactions and takes place in the perceptual organization leading to the motion percept. Another study by Donohue et al. (2015) sought to determine the influence of spatial attention on the temporal window of audiovisual interactions and binding. The experimental design was based on the stream/bounce illusion, in which the timing of a static click can lead to two moving visual objects either streaming

through each other or bouncing off each other. The categorization of moving objects (stream vs. bounce) was dependent on the onset timing between the sound and the intersection of moving objects, which is also called temporal window of integration. Endogenous visuospatial attention narrowed the temporal window of integration, resulting in a decrease in audiovisual interactions. More importantly, they also examined such effects of spatial attention on the temporal profile/window by changing the perceptual task and stimulation. When the participants reported the simultaneity of click with the intersection of the moving objects, the spatial attention widened the temporal window. On the other hand, there was no effect of attention when the task was to report the simultaneity of the same click with the discrete visual flashes. These results revealed the flexible use of attention for audiovisual interactions and associated processes by indicating that the influences of spatial attention are dependent on the stimulus complexity and task demands. Given that speed judgment requires different criterion content than motion direction and categorization (e.g., stream vs. bounce), our results here provide additional evidence for the flexible and adaptive nature of spatial attention.

Manipulation of attentional demands with a secondary task

In the current study, we manipulated attentional demands and perceptual load using a dual-task paradigm. Our results demonstrate that robust auditory time interval effects on perceived speed can be induced even in the presence of a secondary task. Importantly, the amount of these effects was differentially altered when participants performed an additional secondary task. In agreement with the perceptual load theory and previous research (e.g., Alsius et al., 2005), the effects of auditory time interval effects on moving objects decreased when attention was directed to a task-irrelevant stationary visual object (i.e., fixation target). Therefore, these findings point to a significant decrease in the interaction between moving objects and auditory clicks. Previous findings suggest that the origin of such a decrease is mainly due to alterations in the audiovisual binding process (i.e., bimodal processing). However, it is still conceivable that changes in unimodal visual processing may be the origin of the observed decrease in our design. In other words, orienting attention to a task-irrelevant stationary target can suppress visual motion processing and subsequently lead to an overall reduction in audiovisual interactions and auditory time interval effects on perceived speed. It is also important to note that based on the optimal combination of visual and auditory signals (Alais & Burr, 2004), suppression of visual motion signals (a decrease in the quality of motion signals) may lead to an increase in auditory time interval effects on perceived speed. The absence of visual-only (i.e., unimodal) conditions in our design and a behavioral

measure based on the speed comparison performance do not allow us to evaluate the contribution of these alternative accounts directly.

We found that a secondary task on sound location increased the temporal ventriloquism effects on perceived visual speed. Thus, these findings suggest that a focus of attention on the auditory domain can facilitate audiovisual interactions in time. Also, the performance on the secondary task significantly decreased when the distance between speakers was reduced. However, the speaker distance did not alter the temporal ventriloquism effects, and the increase in these effects was due to attention to sound location. For temporal ventriloquism and its influences on different aspects of vision, previous evidence strongly suggests that spatial factors in the auditory domain are not very important, if at all. For instance, Vroomen and Keetels (2006) found that the temporal ventriloquist effects were unaffected by whether sounds came from the same or a different position as the lights, or whether they came from the same or opposite sides of fixation. Thus, spatial correspondence (even crude) is not required for this illusion. In support of this conclusion, the temporal ventriloquism effects on perceived speed have been found to exist when auditory clicks are introduced either through headphones (Ogulmus et al., 2018) or speakers (Kafaligonul & Stoner, 2010). Our findings are in line with the general characteristics of the temporal illusion studied here. An explanation for why temporal ventriloquism effects on perceived speed were enhanced can be based on the facilitatory effects of attention on the unimodal processing of auditory stimuli. Orienting attention to the auditory domain has been shown to improve the perception of auditory stimuli (e.g., Spence & Driver, 1994; Tata et al., 2001; Tata & Ward, 2005). Therefore, a focus of attention on auditory clicks via a secondary task may have improved auditory signals and associated processes, thereby increasing the effects of auditory timing on perceived visual speed. In other words, attention may mainly increase the unimodal auditory signals and hence affect audiovisual processing and their influences on perceived visual speed. Alternatively, rather than altering unimodal auditory processing, attention may directly facilitate audiovisual interactions and their effects on perceived visual speed. Future work will be informative to comprehensively evaluate these alternatives and to further understand the effects of attention at different levels of sensory processing.

Number of moving objects

As mentioned above, our results did not reveal consistent effects of the number of moving objects in the visual field. Temporal ventriloquism effects on perceived speed were present in all the conditions and did not decrease when the number of moving objects was increased. In other words, regardless of the number of objects in each motion frame, the time interval

delineated by a static click successfully drove the timing of multiple moving objects, affecting perceived visual speed. Therefore, our results suggest that audiovisual binding in the temporal domain is not restricted to one visual event and a single auditory event. Our findings are rather in line with recent experimental findings and theoretical framework on audiovisual integration. Using a series of experiments, Boyce, Whiteford, et al. (2020b) found that audiovisual interactions in the temporal domain (e.g., temporal ventriloquism) are not strictly limited to feature similarity/crossmodal correspondence. According to the Bayesian framework on multisensory processing (e.g., Körding et al., 2007; Shams, 2012), they further proposed that audiovisual integration takes advantage of evidence from various processes, assigning different weightings to each process based on relative spatial and temporal characteristics, number of stimuli, and featural characteristics. Using a Bayesian integration approach, Chen et al. (2018) also argued that the effects of auditory timing on visual motion perception are mainly predicted by partial-cue integration, taking into account both temporal proximity and similarity. Together with these recent findings and notion, our findings reveal the existence of temporal ventriloquism in complex stimulation profiles and show that the timing of a brief auditory event can alter motion perception in complex visual scenes (Kafaligonul & Stoner, 2012; Kawachi et al., 2014; Ogulmus et al., 2018).

Conclusion

To conclude, our findings provide important insights into the multisensory nature of motion and speed estimation. We found that the timing of a static click can drive the perception of multiple moving objects in a visual display. At the same time, our results revealed an important modulatory role of attentional demands in the visual and auditory domains, illustrating a decrease in the crossmodal interactions with visual attention, in contrast to an increase in the same paradigm with auditory attention. These findings have important implications for speed estimation in daily life situations in which there is often more than one moving object in cluttered scenes and sensory relevance and attentional demands constantly change.

Declarations

Conflicts of interest The authors have no actual or potential conflicts of interest.

Ethical approval All experimental procedures were in accordance with the Declaration of Helsinki and international guidelines and approved by the local ethics committee.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beer, A. L., & Röder, B. (2004a). Attention to motion enhances processing of both visual and auditory stimuli: An event-related potential study. *Cognitive Brain Research, 18*(2), 205–225. https://doi.org/10.1016/j.cogbrainres.2003.10.004

Beer, A. L., & Röder, B. (2004b). Unimodal and crossmodal effects of endogenous attention to visual and auditory motion. *Cognitive, Affective, & Behavioral Neuroscience, 4*(2), 230–240. https://doi.org/10.3758/cabn.4.2.230

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics, 62*(2), 321–332. https://doi.org/10.3758/BF03205552

Boyce, W. P., Lindsay, A., Zgonnikov, A., Rañó, I., & Wong-Lin, K. (2020a). Optimality and limitations of audio-visual integration for cognitive systems. *Frontiers in Robotics and AI, 7*, 94. https://doi.org/10.3389/frobt.2020.00094

Boyce, W. P., Whiteford, S., Curran, W., Freegard, G., & Weidemann, C. T. (2020b). Splitting time: Sound-induced illusory visual temporal fission and fusion. *Journal of Experimental Psychology: Human Perception and Performance, 46*(2), 172–201. https://doi.org/10.1037/xhp0000703

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436. https://doi.org/10.1163/156856897x00357

Burr, D., & Thompson, P. (2011). Motion psychophysics: 1985-2010. *Vision Research, 51*(13), 1431–1456. https://doi.org/10.1016/j.visres.2011.02.008

Burr, D., Banks, M., & Morrone, M. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research, 198*(1), 49–57. https://doi.org/10.1007/s00221-009-1933-z

Chen, L., & Vroomen, J. (2013). Intersensory binding across space and time: A tutorial review. *Attention, Perception, & Psychophysics, 75*(5), 790–811. https://doi.org/10.3758/s13414-013-0475-4

Chen, L., Zhou, X., Müller, H. J., & Shi, Z. (2018). What you see depends on what you hear: Temporal averaging and crossmodal integration. *Journal of Experimental Psychology: General, 147*(12), 1851–1864. https://doi.org/10.1037/xge0000487

Donohue, S. E., Green, J. J., & Woldorff, M. G. (2015). The effects of attention on the temporal integration of multisensory stimuli. *Frontiers in Integrative Neuroscience, 9*, 32. https://doi.org/10.3389/fnint.2015.00032

Elkin, L. A., Kay, M., Higgins, J. J., & Wobbrock, J. O. (2021). An aligned rank transform procedure for multifactor contrast tests. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '21)* (pp. 754–768). ACM Press. https://doi.org/10.1145/3472749.3474784

Eramudugolla, R., Kamke, M. R., Soto-Faraco, S., & Mattingley, J. B. (2011). Perceptual load influences auditory space perception in the ventriloquist aftereffect. *Cognition, 118*(1), 62–74. https://doi.org/10.1016/j.cognition.2010.09.009

Evans, K. K. (2020). The role of selective attention in cross-modal interactions between auditory and visual features. *Cognition, 196*, Article 104119. https://doi.org/10.1016/j.cognition.2019.104119

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics, 63*(4), 719–725. https://doi.org/10.3758/bf03194432

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software, 8*(15), 1–27. https://doi.org/10.18637/jss.v008.i15

Freeman, E., & Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. *Current Biology, 18*(16), 1262–1266. https://doi.org/10.1016/j.cub.2008.07.066

Getzmann, S. (2007). The effect of brief auditory stimuli on visual apparent motion. *Perception, 36*(7), 1089–1103. https://doi.org/10.1068/p5741

Higgins, J. J., & Tashtoush, S. (1994). An aligned rank transform test for interaction. *Nonlinear World, 1*(2), 201–211.

Higgins, J. J., Blair, R. C., & Tashtoush, S. (1990). The aligned rank transform procedure. *Proceedings of the Conference on Applied Statistics in Agriculture* (pp. 185–195). https://doi.org/10.4148/2475-7772.1443

Kafaligonul, H., & Stoner, G. R. (2010). Auditory modulation of visual apparent motion with short spatial and temporal intervals. *Journal of Vision, 10*(12), Article 31. https://doi.org/10.1167/10.12.31

Kafaligonul, H., & Stoner, G. R. (2012). Static sound timing alters sensitivity to low-level visual motion. *Journal of Vision, 12*(11), 2. https://doi.org/10.1167/12.11.2

Kawachi, Y., Grove, P. M., & Sakurai, K. (2014). A single auditory tone alters the perception of multiple visual events. *Journal of Vision, 14*(8), Article 16. https://doi.org/10.1167/14.8.16

Kaya, U., & Kafaligonul, H. (2019). Cortical processes underlying the effects of static sound timing on perceived visual speed. *NeuroImage, 199*, 194–205. https://doi.org/10.1016/j.neuroimage.2019.05.062

Kaya, U., Yildirim, F. Z., & Kafaligonul, H. (2017). The involvement of centralized and distributed processes in sub-second time interval adaptation: An ERP investigation of apparent motion. *European Journal of Neuroscience, 46*(8), 2325–2338. https://doi.org/10.1111/ejn.13691

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception, 36*(ECVP Abstract Supplement), 14.

Klimova, M., Nishida, S., & Roseboom, W. (2017). Grouping by feature of cross-modal flankers in temporal ventriloquism. *Scientific Reports, 7*, Article 7615. https://doi.org/10.1038/s41598-017-06550-z

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychologica, 134*(3), 372–384. https://doi.org/10.1016/j.actpsy.2010.03.010

Kolers, P. A. (1972). *Aspects of motion perception*. Pergamon Press.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception.

PLOS ONE, 2(9), Article e943. https://doi.org/10.1371/journal.pone.0000943

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

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

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17*(1), 154–163. https://doi.org/10.1016/S0926-6410(03)00089-2

Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2008). Modality-specific selective attention attenuates multisensory integration. *Experimental Brain Research, 184*(1), 39–52. https://doi.org/10.1007/s00221-007-1080-3

Nakayama, K. (1985). Biological image motion processing: A review. *Vision Research, 25*(5), 625–660. https://doi.org/10.1016/0042-6989(85)90171-3

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research, 29*(11), 1631–1647. https://doi.org/10.1016/0042-6989(89)90144-2

Nishida, S. (2011). Advancement of motion psychophysics: Review 2001–2010. *Journal of Vision, 11*(5), Article 11. https://doi.org/10.1167/11.5.11

Ogulmus, C., Karacaoglu, M., & Kafaligonul, H. (2018). Temporal ventriloquism along the path of apparent motion: Speed perception under different spatial grouping principles. *Experimental Brain Research, 236*(3), 629–643. https://doi.org/10.1007/s00221-017-5159-1

Olivers, C. N. L., Awh, E., & Van der Burg, E. (2016). The capacity to detect synchronous audiovisual events is severely limited: Evidence from mixture modeling. *Journal of Experimental Psychology: Human Perception and Performance, 42*(12), 2115–2124. https://doi.org/10.1037/xhp0000268

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437–442. https://doi.org/10.1163/156856897x00366

R Core Team. (2021). *R: A language and environment for statistical computing [Computer software]*. R Foundation for Statistical Computing https://www.r-project.org/

Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology, 89*(2), 1078–1093. https://doi.org/10.1152/jn.00706.2002

Ren, Y., Li, S., Wang, T., & Yang, W. (2020). Age-related shifts in theta oscillatory activity during audiovisual integration regardless of visual attentional load. *Frontiers in Aging. Neuroscience, 12*, Article 571950. https://doi.org/10.3389/fnagi.2020.571950

Ren, Y., Zhao, N., Li, J., Bi, J., Wang, T., & Yang, W. (2021). Auditory attentional load modulates audiovisual integration during auditory/visual discrimination. *Advances in Cognitive Psychology, 17*(3), 193–202. https://doi.org/10.5709/acp-0328-0

Salter, K. C., & Fawcett, R. F. (1993). The art test of interaction: A robust and powerful rank test of interaction in factorial models. *Communications in Statistics: Simulation and Computation, 22*(1), 137–153.

Sanabria, D., Soto-Faraco, S., & Spence, C. (2007). Spatial attention and audiovisual interactions in apparent motion. *Journal of Experimental Psychology: Human Perception and Performance, 33*(4), 927–937. https://doi.org/10.1037/0096-1523.33.4.927

Shams, L. (2012). Early integration and Bayesian causal inference in multisensory perception. In M. M. Murray & M. T. Wallace (Eds.), *Frontiers in the neural bases of multisensory processes* (pp. 217–231). CRC Press.

Shi, Z., Chen, L., & Müller, H. J. (2010). Auditory temporal modulation of the visual Ternus effect: The influence of time interval. *Experimental Brain Research, 203*(4), 723–735. https://doi.org/10.1007/s00221-010-2286-3

Soto-Faraco, S., & Väljamäe, A. (2012). Multisensory interactions during motion perception. In M. M. Murray & M. T. Wallace (Eds.), *The neural bases of multisensory processes* (pp. 579–594). CRC Press.

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., & Kingstone, A. (2002). The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities. *Cognitive Brain Research, 14*(1), 139–146. https://doi.org/10.1016/S0926-6410(02)00068-X

Soto-Faraco, S., Kingstone, A., & Spence, C. (2003). Multisensory contributions to the perception of motion. *Neuropsychologia, 41*(13), 1847–1862. https://doi.org/10.1016/S0028-3932(03)00185-4

Spence, C. J., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance, 20*(3), 555–574. https://doi.org/10.1037/0096-1523.20.3.555

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), 1005–1030. https://doi.org/10.1037/0096-1523.22.4.1005

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

Tata, M. S., & Ward, L. M. (2005). Spatial attention modulates activity in a posterior "where" auditory pathway. *Neuropsychologia, 43*(4), 509–516. https://doi.org/10.1016/j.neuropsychologia.2004.07.019

Tata, M. S., Prime, D. J., McDonald, J. J., & Ward, L. M. (2001). Transient spatial attention modulates distinct components of the auditory ERP. *NeuroReport, 12*(17), 3679–3682. https://doi.org/10.1097/00001756-200112040-00015

Teder-Sälejärvi, W. A., Münte, T. F., Sperlich, F., & Hillyard, S. A. (1999). Intra-modal and cross-modal spatial attention to auditory and visual stimuli. An event-related brain potential study. *Cognitive Brain Research, 8*(3), 327–343. https://doi.org/10.1016/s0926-6410(99)00037-3

Theeuwes, J., & Failing, M. (2020). *Attentional selection: Top-down, bottom-up and history based biases (Elements in Perception)*. Cambridge University Press. https://doi.org/10.1017/9781108891288

Van der Burg, E., Awh, E., & Olivers, C. N. L. (2013). The capacity of audiovisual integration is limited to one item. *Psychological Science, 24*(3), 345–351. https://doi.org/10.1177/0956797612452865

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: A tutorial review. *Attention, Perception, & Psychophysics, 72*(4), 871–884. https://doi.org/10.3758/app.72.4.871

Vroomen, J., Bertelson, P., & de Gelder, B. (2001a). Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychologica, 108*(1), 21–33. https://doi.org/10.1016/S0001-6918(00)00068-8

Vroomen, J., Bertelson, P., & de Gelder, B. (2001b). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics, 63*(4), 651–659. https://doi.org/10.3758/BF03194427

Vroomen, J., & Keetels, M. (2006). The spatial constraint in intersensory pairing: No role in temporal ventriloquism. *Journal of Experimental Psychology: Human Perception and Performance, 32*(4), 1063–1071. https://doi.org/10.1037/0096-1523.32.4.1063

Ward, L. M. (2008). Attention. *Scholarpedia, 3*(10), Article 1538.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*(3), 638–667. https://doi.org/10.1037/0033-2909.88.3.638

Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling and goodness-of-fit. *Perception & Psychophysics, 63*(8), 1293–1313. https://doi.org/10.3758/BF03194544

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics, 63*(8), 1314–1329. https://doi.org/10.3758/BF03194545

Wilbiks, J. M. P., & Dyson, B. J. (2016). The dynamics and neural correlates of audio–visual integration capacity as determined by temporal unpredictability, proactive interference, and SOA. *PLOS ONE, 11*(12), Article e0168304. https://doi.org/10.1371/journal.pone.0168304

Wilbiks, J. M. P., & Dyson, B. J. (2018). The contribution of perceptual factors and training on varying audiovisual integration capacity. *Journal of Experimental Psychology: Human Perception and Performance, 44*(6), 871–884. https://doi.org/10.1037/xhp0000503

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only ANOVA procedures. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11)* (pp. 143–146). ACM Press. https://doi.org/10.1145/1978942.1978963

World Medical Association. (2013). Declaration of Helsinki: Ethical principles for medical research involving human subjects. *Journal of the American Medical Association, 310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution, 1*(1), 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x