Experimental Brain Research (2018) 236:629–643
https://doi.org/10.1007/s00221-017-5159-1

RESEARCH ARTICLE

Temporal ventriloquism along the path of apparent motion: speed perception under different spatial grouping principles

Cansu Ogulmus[1,2,3] · Merve Karacaoglu[1,3] · Hulusi Kafaligonul[1,2,4]

Received: 29 May 2017 / Accepted: 18 December 2017 / Published online: 28 December 2017

Abstract

The coordination of intramodal perceptual grouping and crossmodal interactions plays a critical role in constructing coherent multisensory percepts. However, the basic principles underlying such coordinating mechanisms still remain unclear. By taking advantage of an illusion called temporal ventriloquism and its influences on perceived speed, we investigated how audiovisual interactions in time are modulated by the spatial grouping principles of vision. In our experiments, we manipulated the spatial grouping principles of proximity, uniform connectedness, and similarity/common fate in apparent motion displays. Observers compared the speed of apparent motions across different sound timing conditions. Our results revealed that the effects of sound timing (i.e., temporal ventriloquism effects) on perceived speed also existed in visual displays containing more than one object and were modulated by different spatial grouping principles. In particular, uniform connectedness was found to modulate these audiovisual interactions in time. The effect of sound timing on perceived speed was smaller when horizontal connecting bars were introduced along the path of apparent motion. When the objects in each apparent motion frame were not connected or connected with vertical bars, the sound timing was more influential compared to the horizontal bar conditions. Overall, our findings here suggest that the effects of sound timing on perceived speed exist in different spatial configurations and can be modulated by certain intramodal spatial grouping principles such as uniform connectedness.

Keywords Temporal ventriloquism · Apparent motion · Speed perception · Spatial grouping · Audiovisual interactions · Multisensory

Introduction

The integration of information from different senses is central to our perception of the external world. To construct a coherent representation, the brain organizes (i.e., combines and/or segregates) complex streams of information provided by different modalities. Therefore, it is essential to have brain mechanisms coordinating intramodal and crossmodal perceptual grouping to transform multisensory information into coherent percepts. There has been a growing interest in understanding the principles underlying these coordinating mechanisms (for reviews, see Kubovy and Yu 2012; Spence 2015). Audiovisual paradigms have been frequently applied towards understanding this basic issue of multisensory research. Of particular interest is the phenomenon known as temporal ventriloquism, in which the timing of brief sounds drives the perceived timing of visual events (Fendrich and Corballis 2001; Morein-Zamir et al. 2003; Recanzone 2003). Temporal ventriloquism has been mostly demonstrated using pairs of brief auditory and visual stimuli. In a typical demonstration, the perceived time interval between two visual flashes is altered by the time interval between two auditory clicks. Temporal ventriloquism has been mostly described as brief sounds capturing the timing of visual events in time. Moreover, the capturing effects of brief sounds are explained by the superior temporal resolution of the auditory system (Burr et al. 2009; Welch and Warren 1980). Keetels et al. (2007) examined the influence of auditory (intramodal) grouping on temporal ventriloquism

✉ Hulusi Kafaligonul
hulusi@bilkent.edu.tr

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

2 Interdisciplinary Neuroscience Program, Bilkent University, Ankara, Turkey

3 Department of Psychology, Bilkent University, Ankara, Turkey

4 Aysel Sabuncu Brain Research Center, Bilkent University, 06800 Ankara, Turkey

(crossmodal interactions in time). In addition to capturing sounds, they also used a sequence of background sounds in their experiments. When the capturing sounds had the same features as the background sounds, they grouped together with the background stimulation in time, and hence, the temporal ventriloquism effects became absent. The crossmodal illusion of intersensory pairing in time was only present when the capturing sounds were perceived as segregated from the background sequence. These findings, taken with findings from other similar studies (see also Sanabria et al. 2004; Watanabe and Shimojo 2001), suggest that intramodal grouping has priority over intersensory pairing (and/or crossmodal interactions).

Temporal ventriloquism has also been studied in other aspects of vision which heavily depend on the timing of visual stimulation. For instance, crossmodal temporal capture has been found to alter the perception of visual apparent motion, a motion type induced by rapid presentation and grouping of static motion frames (Getzmann 2007; Kafaligonul and Stoner 2012). The timing of brief sounds can modulate the perceived direction (Freeman and Driver 2008) and speed (Kafaligonul and Stoner 2010) of apparent motion. Moreover, it was shown that sound timing can influence Ternus illusion, in which the timing of apparent motion frames plays a critical role in grouping of visual objects (Ternus 1926, 1938). In a typical Ternus apparent motion display, two motion frames are used and each motion frame consists of vertically aligned objects such as discs. In the second motion frame, these aligned objects are displaced to the right, such that the leftmost object from the first frame appears at the location originally occupied by its neighbor in the first frame. Based on the timing between the motion frames, observers perceive either only the outermost object moving (element motion) or all the objects moving together (group motion) from left-to-right. Using Ternus diplays, Shi et al. (2010) demonstrated that sound timing can shift the percept from element to group motion, and vice versa. Such effects of sound timing on apparent motion perception have been mostly interpreted from the perspective of temporal ventriloquism. The brief sounds capture each motion frame in time and drive the timing of these visual events (or the time interval between these events). Accordingly, this alters the spatiotemporal grouping of motion frames which yields the final illusory perception of motion. Therefore, these temporal ventriloquism studies on apparent motion provide evidence that audiovisual interactions (intersensory pairing) in time can also modulate the percepts that depend on intramodal grouping in vision.

Apparent motion provides an ideal framework for investigating the principles underlying the coordination of intramodal grouping and crossmodal interactions. In the current study, our primary goal was to identify the relationship between spatial (intramodal) grouping of vision and audiovisual interactions in time. Using the temporal ventriloquism influences on the perceived speed of apparent motion (Kafaligonul and Stoner 2010), we tested the basic hypothesis of whether the spatial grouping of vision precedes audiovisual pairing in time. In our experiments, we used visual stimuli containing motion frames including more than one object and manipulated the spatial distance, orientation, and connectedness among these objects. The manipulations on distance and connectedness engaged the proximity and uniform connectedness principles in space, respectively. A change in the orientation led to moving objects in different directions and hence mostly affected the similarity and common fate among these objects. To quantify the audiovisual interactions in time, observers compared the speed of an apparent motion display with a distinct spatial configuration across different sound timing conditions. When a particular Gestalt grouping principle (e.g., spatial proximity) was prior to audiovisual interactions in time, we expected that the specific manipulation (e.g., spatial distance) of that grouping principle would significantly modulate the temporal capture effect of brief sounds and the perceived speed difference between different sound timing conditions. Moreover, there is an increasing interest in understanding the limiting factors of audiovisual integration in complex visual configurations (Takeshima and Gyoba 2013; Van der Burg et al. 2014). To date, temporal ventriloquism has not been investigated systematically using different complex visual displays. Our experiments here included displays with varying numbers of moving objects, and the spatial relationships between these objects were changed systematically across conditions. Therefore, within the context of apparent motion, our findings here also contribute to characterizing audiovisual temporal processing in complicated visual scenes.

Experiment 1a

In the first experiment, we examined the temporal ventriloquism effects on perceived speed under different Gestalt spatial grouping principles. In each motion frame, we mainly used two flashed bars and manipulated the spatial distance, orientation, and connectedness between these visual objects.

Methods

Participants

Fourteen naïve observers (age range 21–33 years) participated in this experiment. In all of the experiments, observers had normal hearing, and normal or corrected-to-normal visual acuity. Participants gave informed consent prior to the experimental sessions and all procedures were in accordance with international standards (Declaration of Helsinki,

1964) and approved by the ethics committee at Faculty of Medicine, Ankara University.

Apparatus

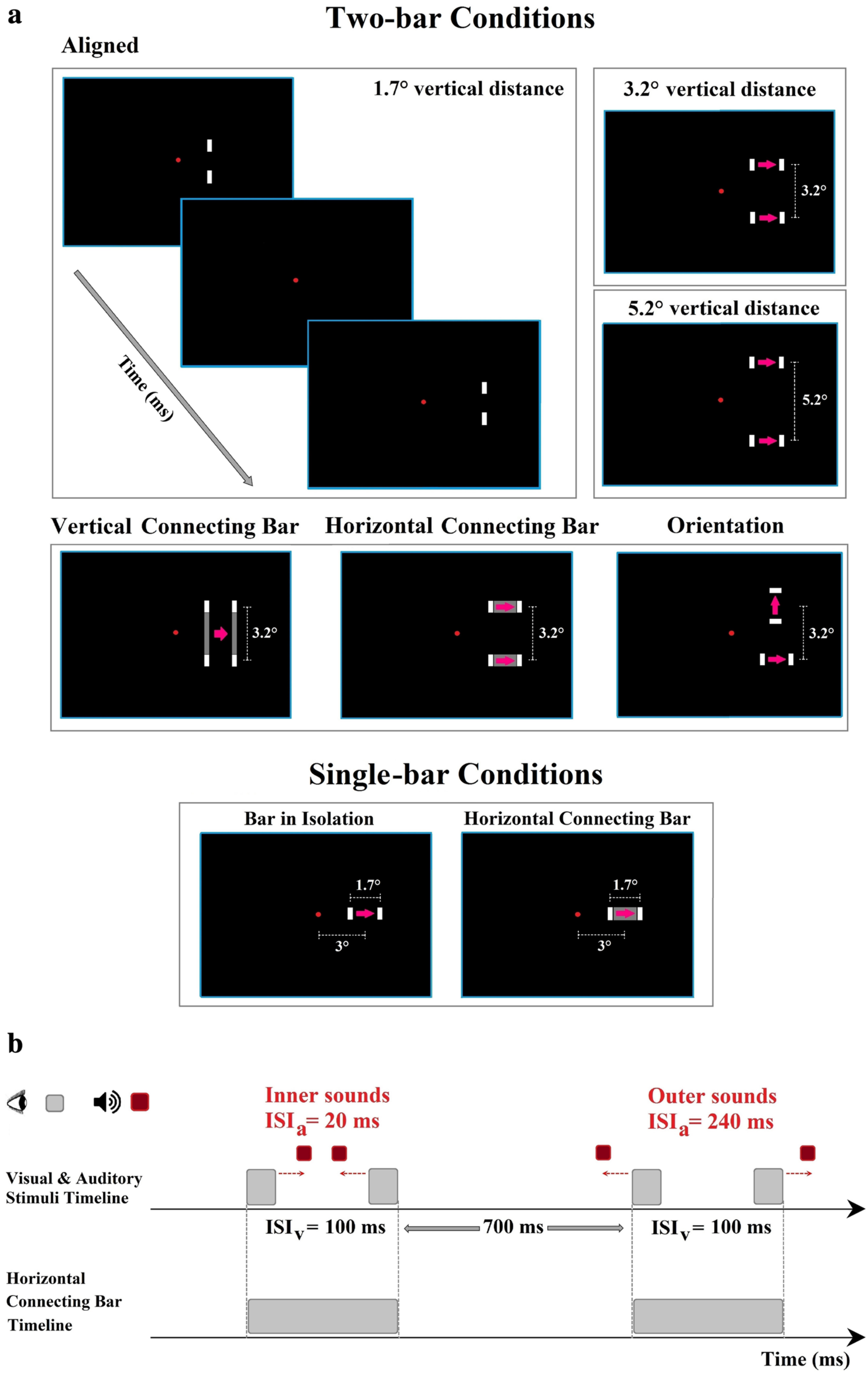
We used Matlab version 7.12 (The MathWorks, Natick, MA) with Psychtoolbox 3.0 (Brainard 1997; Pelli 1997) for stimulus presentation and data acquisition. Visual stimuli were presented on a 20-inch CRT monitor (HP p1230, 1280 × 1024 pixel resolution and 100 Hz refresh rate) at a viewing distance of 57 cm. A SpectroCAL photometer (Cambridge Research Systems, Rochester, Kent, UK) was used for luminance calibration and gamma correction of the display. Auditory stimuli were introduced via headphones (Sennheiser HD518, Sennheiser Electronic GmbH & Co. KG, Germany) and amplitudes were measured by a sound-level meter (SL-4010 Lutron, Lutron Electronics, Taipei, TW). Timing of auditory and visual stimuli was confirmed with a digital oscilloscope (Rigol DS 10204B, GmbH, Puchheim, Germany) connected to the computer soundcard and a photodiode (which detected visual stimulus onsets). Head movements were constrained by a chin rest. The experiments were performed in a dark, sound-attenuated psychophysics room. The same apparatus and testing room were used in all the experiments.

Stimuli and procedure

A small red circle (0.25° diameter) at the center of the display served as a fixation point. Visual stimuli consisted of two apparent motion frames (Fig. 1a). In each motion frame, either a single or two bars (0.3° × 1.2° with a luminance of 28.45 cd/m^2) were presented on a black background (0.36 cd/m^2). The duration of each motion frame was 50 ms, and the blank time interval (inter-stimulus interval, visual ISI_v), during which only the fixation point was present, between each motion frame was 100 ms. When a single bar was presented in each motion frame (single bar in isolation), the bar locations were adjusted in the right visual field, such that the center of the apparent motion was located 3° away from the fixation circle in the horizontal dimension. The horizontal distance (i.e., center-to-center separation between left and right bars) between each flashed bar of the apparent motion was 1.7° and the motion direction was either leftward or rightward. For instance, a percept of rightward movement was generated by flashing a single bar at 2.15° to the right of fixation for 50 ms, then was extinguished for a 100 ms ISI, and then flashing the bar again at 3.85° to the right of fixation. Moreover, these spatial and temporal parameters also led to smooth and continuous apparent motion percept in the other experimental conditions studied in the present study (for definitions and classification of apparent motion, see Ekroll et al. 2008; Getzmann 2007; Kolers 1972). To engage connectedness between the flashed bars of the apparent motion (single horizontal connecting bar condition), we introduced a horizontal gray (2.72 cd/m^2) bar (rectangle). The gray bar was centered with the apparent motion and it covered the blank space between the flashed bars (1.4° × 1.2°) during the presentation of apparent motion (2 × 50 ms motion frame duration + 100 ms ISI, see also Fig. 1b for a timeline). In other words, the connectedness between the start and end points of apparent motion was introduced presenting a horizontal connecting bar between these points. The perceived motion is typically altered by such gray bar or rectangle introduced along the path of apparent motion. These changes in illusory motion percept have also been known as "path-guided apparent motion" (Shepard and Zare 1983).

In case of two flashed bars in each motion frame, the vertical locations of bars in the right visual field were symmetrically arranged relative to (above and below) the fixation circle (two-aligned bar condition). To manipulate the spatial proximity between the flashed bars, the vertical distance (i.e., center-to-center separation) between them was pseudo-randomly selected from three values: 1.7, 3.2, and 5.2°. It should be noted that the spatial gap between the two bars (0.5°) was smaller than the horizontal gap (1.4°) along the motion direction for the 1.7° vertical distance condition, whereas it was larger for the other vertical distance conditions. In all the two-aligned bar conditions, observers typically perceived more than one moving object in the visual display. All other parameters of flashed bars were the same as those of the single-bar conditions. As shown in Fig. 1a, vertical and horizontal connecting gray (2.72 cd/m^2) bars were used to engage connectedness. In the vertical connecting bar condition, the gray bars were placed within the vertical space between the bars of each motion frame. For the horizontal connecting bar condition, we presented a gray horizontal rectangle (1.4° × 1.2°) along the path of each apparent motion (i.e., in the center of two sequentially flashed bars). Compared to vertical connecting bar condition, observers typically perceived more than one moving object in the horizontal connecting bar condition. We manipulated the similarity and common fate by rotating the upper flashed bar 90° (orientation condition). This led to two simultaneous apparent motions in different directions. More specifically, the upper one was perceived as moving in the vertical (upward or downward) direction, while the lower one was perceived as moving in the horizontal (leftward or rightward) direction. The vertical distances between these two motions were the same as those (1.7, 3.2, or 5.2°) used in the other two-bar conditions.

Auditory stimuli were two 20 ms static clicks, comprised of a rectangular windowed 480 Hz sine-wave carrier, sampled at 44.1 kHz with 8-bit quantization. They were introduced binaurally at 83 dB sound pressure level



(SPL). For each trial, the visual displays in the apparent motion frames were pseudo-randomly selected from 14 (3 vertical distances × 4 two-bar configurations and 2 single-bar) conditions. As shown in Fig. 1b, during each trial, the two-frame apparent motion stimuli were shown twice with a temporal delay of 700 ms. Each consecutive presentation was exactly the same in terms of visual stimulation, but the timing of the auditory stimuli differed. A pair of clicks was centered temporally within each presentation of two-frame apparent motion and the time interval (auditory ISI_a)

◂**Fig. 1** **a** Spatial configurations of the apparent motion conditions used in Experiment 1a. Each apparent motion consisted of two motion frames. In each frame, there were either a single or two flashed bars. In the case of two flashed bars, the vertical distance (spatial proximity) between the bars was changed across conditions. The connectedness (vertical vs. horizontal connecting bar) between these bars and the orientation of the upper bar (aligned vs. orientation) were also manipulated. All the vertical distance (1.7, 3.2 and 5.2°) conditions of two-aligned apparent motions are shown, and only the 3.2° vertical distance conditions of the other configurations are displayed. The apparent motion displays including a single bar (with or without horizontal connecting bar) are shown at the bottom. In all the conditions, we used the same horizontal eccentricity and separation between the flashed bars. To avoid clutter, these values are only marked on the single-bar conditions. The arrows highlight the apparent motion paths and they were not present in the actual stimuli. **b** Timing diagram of each trial. For each trial, one of the apparent motion displays was selected pseudo-randomly and presented to the observer twice with a temporal delay of 700 ms. In terms of visual stimulation, these consecutive apparent motions were exactly the same. However, the time interval (ISI_a) between static clicks was either shorter (inner sounds) or longer (outer sounds) than the ISI_v between the apparent motion frames. The apparent motion frames and auditory clicks are represented by gray and red squares, respectively. For the horizontal connecting bar conditions, the timing of the horizontal gray rectangle is shown by the diagram at the bottom. Relative durations of visual and auditory events are indicated by thickness of squares and rectangles (height of these icons distinguishes stimulus modality and is otherwise irrelevant)

between them was either 20 ms (inner sounds) or 240 ms (outer sounds). For both sound timing, the ISI between each motion frame and each click (i.e., the first motion frame-the first click or the second motion frame-the second click) was 20 ms. For inner sounds, two-click sequence started after the offset of the first motion frame and ended before the onset of the second motion frame. On the other hand, two clicks of outer sound timing started before the onset of the first motion frame and ended after the offset of the second motion frame. The temporal order of the inner and outer sounds was randomized across trials. At the end of each trial, observers indicated, by pressing one of the two keys, which apparent motion stimuli appeared to move faster (two-interval forced choice). Observers were told that the visual stimuli would be accompanied by clicks but to base their responses solely on the visual stimuli. They were also asked to distribute their attention to the right visual field and to compare speed based on the overall motion in the display when more than one moving object was presented simultaneously.

Each spatial condition was repeated 28 times per session. Accordingly, there were 392 trials (14 spatial conditions × 28 trials per stimulus) in each experimental session. Each participant completed one experimental session. Prior to this experimental session, each participant was also shown examples of the apparent motion stimuli, followed by a practice session without any sound or feedback.

Results and discussion

In agreement with the previous findings, the time interval demarcated by brief sounds affected the perceived speed of two-frame apparent motion. Even though observers compared the speed of physically identical visual motion stimuli, the apparent motion with inner sounds was perceived to move faster than the one with outer sounds in more than 60% of the trials (Fig. 2). If a specific spatial grouping principle precedes the audiovisual interactions in time, then we would expect a significant modulation in the observed effects of sound timing when a change is introduced in that grouping principle. To test this basic prediction, we performed a two-way repeated-measures ANOVA with spatial distance (proximity) and configuration as factors on the data shown in Fig. 2 (left panel). The main effect of spatial distance between two simultaneous apparent motions was not significant ($F_{2,26} = 1.83, p = 0.181, \eta_p^2 = 0.123$). However, the effect of spatial configuration ($F_{3,39} = 12.31, p < 0.001, \eta_p^2 = 0.486$) and its interaction with distance were found to be significant ($F_{6,78} = 3.38$, $p = 0.005$, $\eta_p^2 = 0.206$). To disentangle the sources of this interaction, we performed additional ANOVA tests. A two-way repeated-measures ANOVA with factors connecting bar type (vertical connecting bar vs. horizontal connecting bar) and spatial distance as factors revealed only a significant main effect of connecting bar ($F_{1,13} = 18.08$, $p < 0.001$, $\eta_p^2 = 0.583$). The effect of spatial distance ($F_{2,26} = 1.33, p = 0.281, \eta_p^2 = 0.093$) and two-way interaction ($F_{2,26} = 2.84$, $p = 0.077$, $\eta_p^2 = 0.179$) were not significant. Overall, compared to the vertical connecting bar condition, the perceived speed difference between the two sound timing was smaller for the horizontal connecting bar condition. We also applied a similar approach to assess the effect of orientation (i.e., motion direction). The ANOVA with factors orientation (aligned vs. orientation) and spatial distance revealed only a significant effect of spatial distance ($F_{2,26} = 4.22, p = 0.026, \eta_p^2 = 0.245$). On the other hand, the effect of whether two apparent motions have the same orientation (i.e., move in the same direction) or not ($F_{1,13}$ = 1.24, p = 0.286, $\eta_p^2 = 0.123$), and the two-way interaction were not significant ($F_{2,26}$ = 3.19, $p = 0.058, \eta_p^2 = 0.123$). We also performed one-way repeated-measures ANOVAs on each spatial configuration in Fig. 2 (left panel), separately. The main effect of spatial distance was found to be significant only for the orientation condition ($F_{2,26}$ = 5.60, $p = 0.010, \eta_p^2 = 0.300$). When the two bars moved in different directions, the effect of sound timing became smaller as the spatial distance between them was increased.

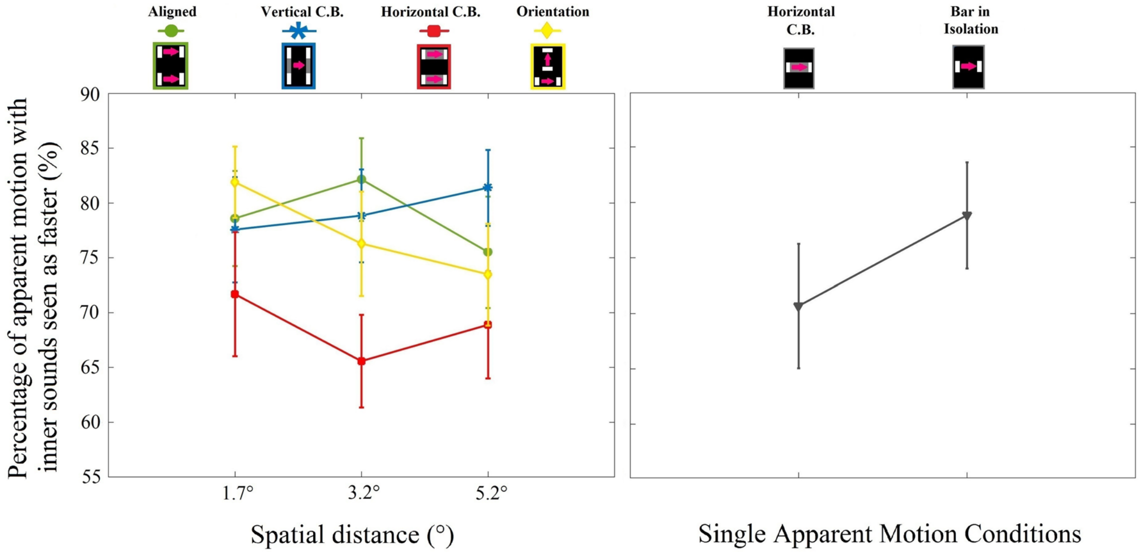


Fig. 2 Results of Experiment 1a ($n=14$). The results of the single-bar and two-bar apparent motion conditions are displayed in separate plots. In the left plot, the percentage of apparent motion with inner sounds seen as faster is displayed as a function of the vertical distance for each of the two-bar configurations. The plot on the right indicates these percentage values for the single-bar conditions. Error bars correspond to ± SEM

Figure 2 (right panel) also shows the results for the single-bar conditions. In line with the results of two-bar conditions, the effect of sound timing was smaller for the horizontal connecting bar condition. In other words, compared to the single bar in isolation (i.e., no horizontal connecting bar) condition, the proportion of trials in which the apparent motion with inner sounds was perceived as faster significantly decreased when there was a faint rectangle flashed during the apparent motion (paired samples two-tailed t test, $t_{13}=3.16$, $p=0.007$). Moreover, we compared the results of single- and two-bar conditions. Bonferroni-corrected pairwise comparisons indicated that the results of single bar in isolation were not significantly different than any two-aligned, vertical connecting and orientation results shown in Fig. 2 (left panel). Similarly, in terms of percentage values, there was no significant difference between the single (with a horizontal connecting bar) and any of two horizontal connecting bar conditions. Overall, these findings suggest that specific spatial grouping principles of vision can take precedence over temporal ventriloquism. In particular, uniform connectedness had consistent effects on temporal ventriloquism and this grouping principle may be prior to temporal ventriloquism.

Experiment 1b

By indicating significant effects of connecting bar type, the results of Experiment 1a highlight the importance of (uniform) connectedness during intersensory pairing in time and temporal ventriloquism. On the other hand, the effects of spatial distance were not consistent across different spatial configurations and the effects of orientation were not significant. In Experiment 1a, we only used two flashed bars in each motion frame (i.e., two simultaneous apparent motions) and the visual stimuli were always presented in the right visual field. Moreover, during the orientation conditions, the observers might have ignored our instructions, attended only to the horizontally moving bar and completely ignored the vertically moving bar. Accordingly, the lack of consistent spatial distance and orientation effects might be due to these limitations in the experimental design. In the present experiment, we re-examined the spatial proximity and similarity/common fate principles in two different visual fields using three simultaneously moving bars.

Methods

Participants

Ten observers (age range 20–30 years) completed this experiment, and all of them were naïve to the purpose of the experiment. Three of these observers took part in Experiment 1a.

Stimuli and procedure

We used three flashed bars in the apparent motion displays. The spatial configurations for the right visual field are shown in Fig. 3 (top panel). In the three-aligned bar condition, three

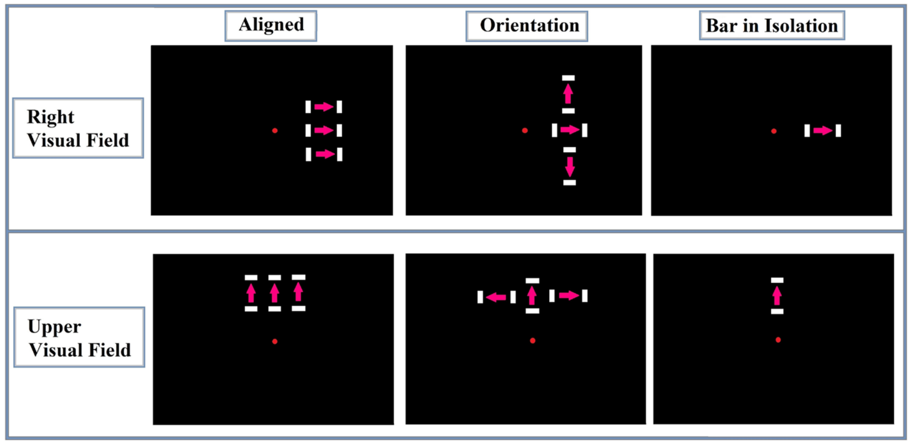


Fig. 3 Spatial configurations of the apparent motion conditions used in Experiment 1b. Each apparent motion frame consisted of either a single or three flashed bars. The conditions for each visual field are shown in separate rows. The arrows highlight the apparent motion paths and they were not present in the actual stimuli. The timing diagrams for each trial were the same as those in Fig. 1b

bars were aligned in each motion frame and this resulted in three simultaneous motion percepts in the same direction (leftward or rightward). The middle apparent motion was vertically centered with the fixation circle and was 3° away from the center of the screen in the horizontal domain. The two surrounding apparent motions were symmetrically arranged relative to the middle one. To generate the orientation condition, the orientations of the surrounding bars were changed, such that the resulting apparent motion directions were perceived in the vertical direction (upward or downward), while the middle one was perceived in the horizontal direction (leftward or rightward). For the upper visual field conditions, we used the same flashed bars (Fig. 3, bottom panel). However, their orientations and locations in the visual field were changed. The center of the middle apparent motion was 3° above the fixation circle. When there was an orientation change, the middle and surrounding bars resulted in motions in the vertical (upward or downward) and horizontal (leftward or rightward) directions, respectively. We manipulated the spatial proximity by changing the vertical and horizontal distances between the apparent motion centers for the right and upper visual fields, respectively. The spatial distance values (1.7, 3.2, or 5.2°) were the same as those used in Experiment 1a. All of these three-bar (i.e., aligned and orientation) conditions typically led to simultaneous and distinct apparent motion percepts in the visual display. For both visual fields, we also included single-bar conditions. In these conditions, we only presented the middle bar and hence only a single motion was perceived.

The right and upper visual field conditions were run in separate experimental sessions. Accordingly, there were 7 (3 spatial distances × 2 three-bar configurations and 1 single-bar) conditions and 196 trials (28 trials per condition) in each experimental session. Each observer completed one experimental session for each visual field and the order of these experimental sessions was randomized across observers. Except for these changes introduced in the visual stimuli and procedure, all other stimulus parameters and experimental procedures were the same as those used in Experiment 1a.

Results and discussion

As in Experiment 1a, for all the conditions studied here, the apparent motion with inner sounds was perceived to be faster than the one with outer sounds in more than 60% of the trials (Fig. 4). To assess these effects of sound timing on perceived speed across different visual displays, we performed a three-way repeated-measures ANOVA with visual field, spatial distance and configuration (aligned vs. orientation) as factors. The main effect of visual field was not significant ($F_{1,9} = 1.107$, $p = 0.320$, $\eta_p^2 = 0.12$). On the other hand, the effects of spatial distance ($F_{2,18} = 4.505$, $p = 0.026$, $\eta_p^2 = 0.33$) and configuration ($F_{1,9} = 23.241$, $p < 0.001$, $\eta_p^2 = 0.72$) were found to be significant. On average, the effect of sound timing was smaller for the 5.2° than the other spatial distances, and the sound timing was more effective for the aligned compared to the orientation conditions. The ANOVA test did not indicate any significant two-way interactions and a three-way interaction between these main factors. Overall, these results indicate significant effects of orientation and proximity in three simultaneously moving bar displays. They suggest that these grouping principles can take precedence over temporal ventriloquism effects on the perceived speed of three simultaneous apparent motions.

In agreement with the ANOVA test results on the three-bar conditions, there was no significant difference between the percentage values of single-bar conditions presented in different visual fields (paired samples two-tailed t test, $t_9 = 0.688$, $p = 0.508$). In addition, we performed pairwise comparisons (Bonferroni-corrected) between the single- and the three-bar conditions of each visual field separately. In terms of percentage values, none of the three-aligned bar conditions were significantly different from the

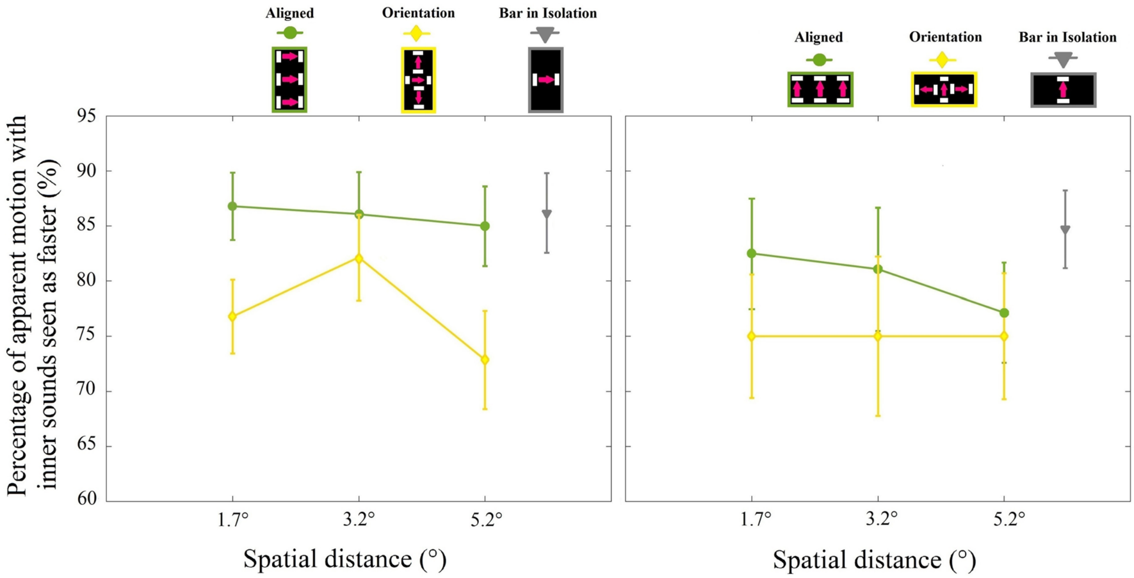


Fig. 4 Results of Experiment 1b ($n = 10$). The results for each visual field are displayed in separate plots (left plot: right visual field, right plot: upper visual field). In each plot, the percentage of apparent motion with inner sounds seen as faster is displayed as a function of the spatial distance. The results for the single apparent motion condition are represented by a separate data point (triangle symbol) in each plot. Error bars correspond to ± SEM

corresponding single-bar condition for either visual field. For the right visual field, the percentages of inner sounds seen as faster for 1.7 and 5.2° orientation conditions were significantly ($p < 0.05$) smaller than that of single-bar condition. Similarly, the result of 1.7° orientation condition was significantly ($p < 0.05$) different than that of the single-bar presented in the upper visual field. These findings do not provide a coherent picture on the relationship between the number of visual items used and the extent temporal ventriloquism effects on perceived speed. Our results only indicate that the effects of sound timing in a few orientation (three-bar) conditions were significantly different than the ones observed in the corresponding single-bar conditions.

The designs of Experiments 1a and 1b were based on judging the speed of consecutive apparent motions with different sound timings. Against our instructions and speed judgement strategy used in the practice sessions (see Experiment 1a procedure), observers could have conceivably ignored apparent motions and relied only on auditory timing for speed judgements in the main experimental sessions. If such response/decisional bias led to the observed differences between inner and outer sounds, the percentage values should not have been affected by low-level stimulus manipulation in vision. However, this is not the case in Experiment 1a and 1b results. For instance, in Experiment 1a, the connectedness between two flashed bars had significant influences on the percentage of trials in which apparent motion inner sounds were perceived as faster. In addition, an orientation change in the surrounding bars had significant effect on the percentage values in three flashed bar configurations (Fig. 4). Accordingly, we consider that the contribution of any response/decisional bias to Experiment 1a and 1b results is unlikely and minimal.

Experiment 2a

In Experiments 1a and 1b, we used a behavioral task that depends on comparing the speed of physically identical visual apparent motions paired with different sound timings. Such a design does not explicitly allow us to measure perceived speed values for different sound timing conditions or how these values might change when visual stimuli are arranged in different spatial configurations. Here, we explicitly quantified perceived speed changes in the temporal domain using a similar approach to that of Kafaligonul and Stoner (2010). As the results of Experiment 1a point to dominant effects of connectedness, we only used the vertical and horizontal connecting bar conditions in three-bar apparent motion displays.

Methods

Participants

Six naïve observers (age range 22–24 years) participated in this experiment. Only one of these observers took part in Experiment 1a.

Stimuli and procedure

As in Experiment 1b, we used three flashed bars (0.3° × 1.2° with a luminance of 39.53 cd/m^2) in each apparent motion frame (Fig. 5a). The middle bar was vertically centered with the fixation circle and the center of the resulting apparent motion was 3° away from the fixation circle in the right visual field. The other bars were arranged 3.2° (center-to-center vertical distance) above and below the middle bar. To engage connectedness between the flashed bars, we used vertical and horizontal connecting gray (1.75 cd/m^2) stimuli. In the vertical connecting bar condition, the gray bars were placed within the space between the flashed bars of each motion frame. For the horizontal connecting bar condition, we presented the gray rectangles (1.4° × 1.2°) along the path of each apparent motion. Each spatial configuration was run in separate experimental sessions.

During each trial, we presented an audiovisual reference stimulus and a silent test stimulus with a temporal offset of 600 ms (Fig. 5b). The reference stimulus had a fixed ISI between each motion frame (ISI_{ref}) of 100 ms and a variable ISI between the clicks (ISI_a). The inner and outer sound conditions had 20 and 240 ms ISI_a, respectively. The *test* included only the visual motion frames and the ISI (ISI_{test}) between them was varied pseudo-randomly from trial to trial: 0, 20, 40, 60, 80, 100, 120, 140, 160, 180, and 200 ms. In all of these ISI_{test} conditions, the test stimulus was mainly rated in the category of continuous and smooth apparent motion. However, the small ISI values led to smoother apparent motion percepts and the average rating values for these conditions were higher than the ones for longer ISI values. At the end of each trial, observers indicated, by pressing one of the two keys, which apparent motion stimulus appeared to move faster (two-interval forced choice). The reference and test moved in the same direction (leftward or rightward). In terms of visual stimulation, the reference and test stimuli were physically identical. They were not distinguished in the instructions to the observers and their temporal order was randomized from trial to trial.

Each experimental session had a balanced mixture of two sound conditions and included a total of 220 trials (2 sound conditions × 11 ISI_{test} × 10 trials per stimulus). Each observer completed four experimental sessions (2 spatial configurations × 2 runs). The order of these sessions was randomized. Prior to these main experimental sessions, each participant was shown examples of the visual apparent motion stimuli (without sounds) followed by two or three practice sessions of visual-only (reference apparent motion without sounds)

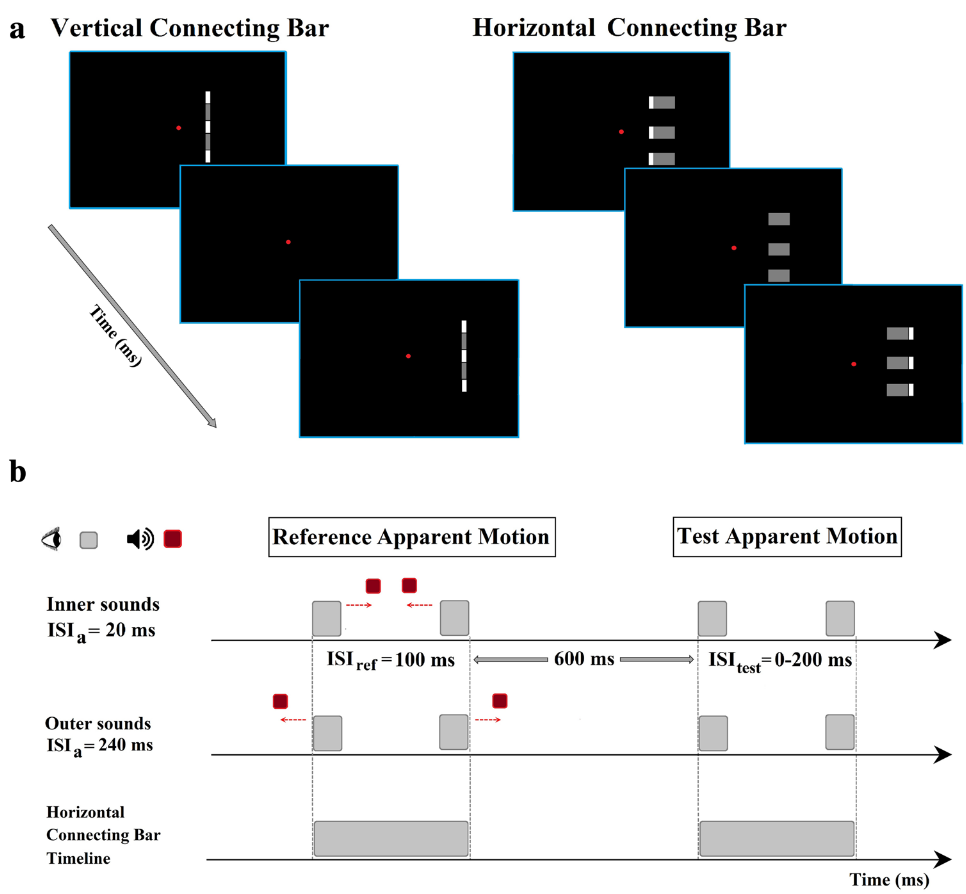


Fig. 5 **a** Spatial configurations of the apparent motion conditions used in Experiment 2a. In each motion frame, there were three bars and the connectedness (vertical vs. horizontal connecting bar) between them was manipulated. **b** Timing diagram for each trial. The reference and test consisted of two apparent motion frames. Auditory clicks were only introduced during the reference stimulus. Each experimental session consisted of inner and outer sound conditions. Other conventions are the same as those in Fig. 1b

trials without any feedback. All other stimulus parameters and experimental procedures were the same as those used in Experiment 1b.

Estimation of perceived speed

For each condition (2 sounds × 2 spatial configurations), a cumulative Gaussian function was fitted to the individual and group-averaged data using psignifit (version 2.5.6), a software package that implements the maximum likelihood method described by Wichmann and Hill (2001a, b). The 50% point on the resultant curves yields the point of subjective equality (PSE). The PSE value is the ISI_{test} for which the test was seen as faster than the reference on 50% of the trials. Thus, by looking for changes in these PSE values, we were able to quantify changes in the perceived speed of the reference stimulus across different experimental conditions.

Results and discussion

If the time interval marked by auditory sounds (auditory ISI_a) does change the perceived visual time interval (ISI_{ref}) and perceived speed, we would expect a systematic change in the PSE values for different auditory time interval conditions. More specifically, for the inner sound condition ($ISI_a < ISI_{ref}$), the ISI_{ref} should appear to contract leading to an increase in the perceived speed of the reference stimulus. On the other hand, for the outer sound condition ($ISI_a > ISI_{ref}$), the ISI_{ref} should appear to lengthen and we expect a decrease in the perceived speed of the reference stimulus. Based on this basic temporal capture account, we predicted the PSE values for inner sound condition to be smaller than those of the outer sound conditions. Group-averaged data shown in Fig. 6a support this prediction. The PSE values for the inner sound conditions were found to be smaller than those for the outer sound conditions. To assess whether the changes in PSE values were significant across different sound timing (inner vs. outer sounds) and spatial configuration (vertical vs. horizontal connecting bar) conditions, we estimated the PSE values from individual observers (Fig. 6b) and performed a two-way repeated-measures ANOVA on these measures. The main effect of sound timing ($F_{1,5} = 545.596$, $p << 0.001$, $\eta_p^2 = 0.99$) and the interaction between sound timing and spatial configuration ($F_{1,5} = 17.821$, $p = 0.008$, $\eta_p^2 = 0.78$) were statistically significant. However, the main effect of spatial configuration was not significant ($F_{1,5} = 0.39$, $p = 0.560$, $\eta_p^2 = 0.07$). To understand the exact nature of the two-way interaction, we performed post hoc pairwise comparisons for each sound condition separately. For the inner sound conditions, the PSE values of the horizontal connecting bars were significantly higher

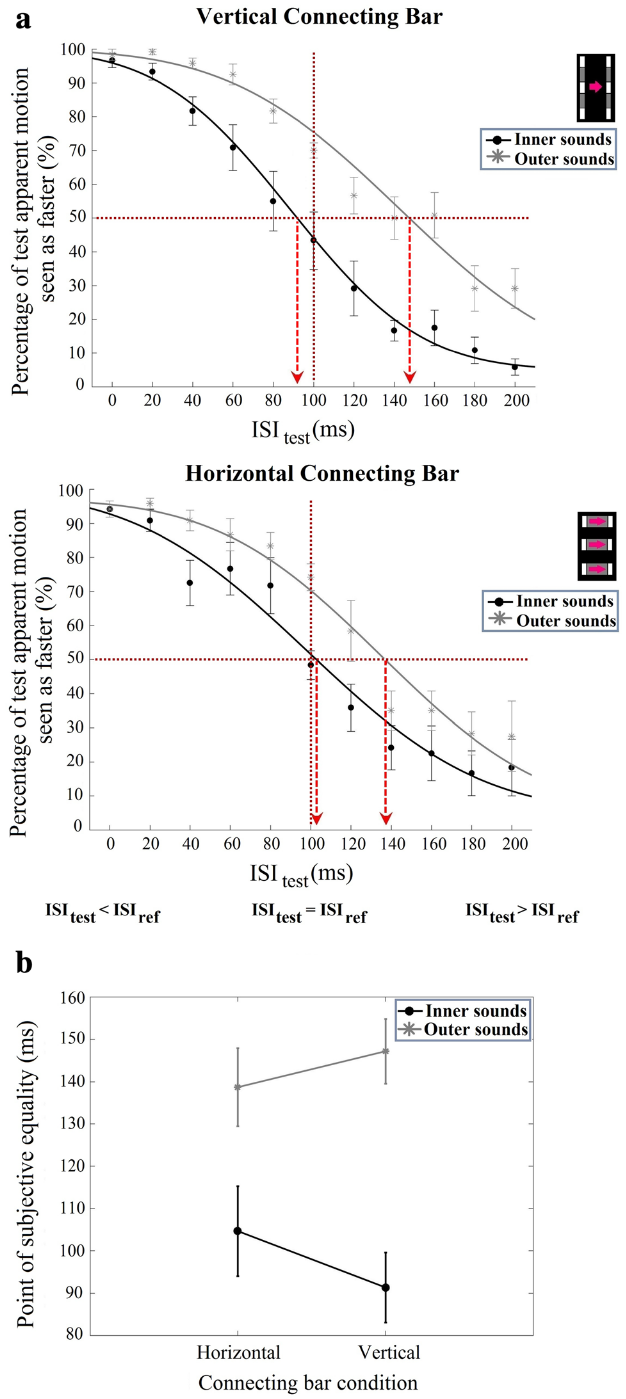


Fig. 6 Results of Experiment 2a ($n = 6$). **a** Group-averaged data for different experimental conditions. Each spatial configuration condition is displayed in separate plots, and in each plot, the inner and outer sound conditions are shown. The plots indicate the percentage of trials in which the test stimulus was judged to move faster than the reference stimulus as a function of ISI_{test}. The black and gray curves correspond to psychometric fits for inner and outer sounds, respectively. The intersection of the 50% point with the vertical line provides an estimate of the PSE for each condition. Error bars correspond to ± SEM. **b** The averaged PSEs of all observers for different experimental conditions. Error bars correspond to ± SEM

than those of the vertical connecting bar conditions (two-tailed, $t_5 = 3.781$, $p = 0.012$). On the other hand, the difference between two spatial configurations was not significantly different in terms of outer sound PSE values (two-tailed, $t_5 = 1.528$, $p = 0.186$). We also computed the difference between the PSE values of the outer and inner sound conditions for each subject. The paired comparison on these difference PSEs (outer–inner) revealed a significant effect of spatial configuration (two-tailed, $t_5 = 4.221$, $p = 0.008$), suggesting that the effect of sound timing (temporal ventriloquism) was significantly smaller for the horizontal condition compared to the vertical connecting bar condition.

We confirmed the significant effects of connectedness in Experiment 1a by another experiment designed to quantify perceived speed changes in the temporal domain. The PSE results here indicate significant effects of both sound timing and connectedness in three-bar apparent motion display. Moreover, the changes in PSE values (i.e., perceived speed) were in agreement with the changes in percentage values shown in Fig. 2 (left panel).

Experiment 2b

In a separate experiment, we also examined how sound timing affected the perceived speed of apparent motion frames containing a single bar. The spatial configurations used in the present experiment resulted in apparent motion percepts with and without horizontal connecting bar.

Methods

Participants

Eight naïve observers (age range 21–25) completed Experiment 2b. Four of these observers took part in Experiment 2a, and two observers took part in Experiment 1a.

Stimuli and procedure

For each apparent motion frame, we used a single bar in the right visual field (Fig. 7). The sequential presentation of this stimulus with a temporal and spatial offset led to an apparent motion 3° away from the fixation circle in the horizontal domain (single bar in isolation). During the presentation of apparent motion (i.e., from the onset of the first motion frame to the offset of the second motion frame), a gray (faint) connecting rectangle was introduced along the path of the apparent motion. This condition was referred to as the single horizontal connecting bar condition. Except for these changes introduced in the visual stimuli, all other stimulus parameters, conditions, and experimental procedures were the same as those used in Experiment 2a.

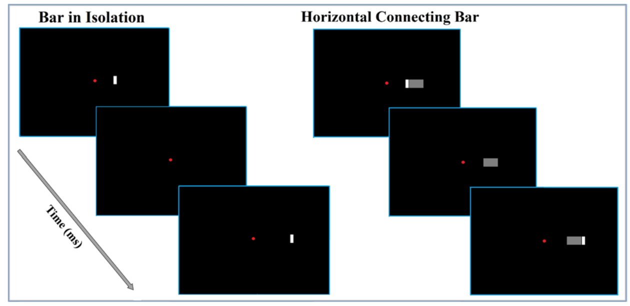


Fig. 7 Timing and spatial configurations for the apparent motion displays used in Experiment 2b. Each motion frame consisted of a single visual bar (with or without horizontal connecting bar). The timing diagrams for each trial were the same as those in Fig. 5b

Results and discussion

The group-averaged data and the estimated PSE values from individual data are shown in Fig. 8. In agreement with the temporal capture account, the PSE values for the inner sound condition were smaller than those for the outer sound conditions (Fig. 8b). As in Experiment 2a, we performed a two-way repeated-measures ANOVA with sound timing (inner vs. outer sounds) and spatial configuration (with vs. without a horizontal connecting bar) as factors. This analysis revealed a significant effect of sound timing ($F_{1,7} = 87.128$, $p < 0.001$, $\eta_p^2 = 0.92$) and a two-way interaction ($F_{1,7} = 9.912$, $p = 0.016$, $\eta_p^2 = 0.59$). The main effect of spatial configuration was not significant ($F_{1,7} = 4.281$, $p = 0.077$, $\eta_p^2 = 0.38$). In the post hoc tests, we compared the PSE values of single-bar conditions for each sound timing condition. The difference between the single bar in isolation and horizontal connecting bar conditions was not significant for the inner sounds (two-tailed, $t_7 = -1.634$, $p = 0.146$). However, in the outer sound conditions, the PSE values of the single bar in isolation were significantly larger than those of the horizontal connecting bar condition (two-tailed, $t_7 = 2.837$, $p = 0.025$). Additional analyses on the difference PSEs (outer–inner) indicated that the difference between the PSE values was significantly larger for the single bar in isolation compared to the horizontal connecting bar condition (two-tailed, $t_7 = 3.148$, $p = 0.016$).

The observed changes on the difference PSEs point to a decrease in the effects of sound timing on perceived speed and temporal ventriloquism when a horizontal gray bar was introduced along the path of apparent motion. By showing PSE changes in the temporal domain, these findings provide more direct evidence that perceived speed is modulated by connectedness and by audiovisual interactions in time via temporal ventriloquism. In this respect, they are in agreement and confirm the observed changes in the percentage

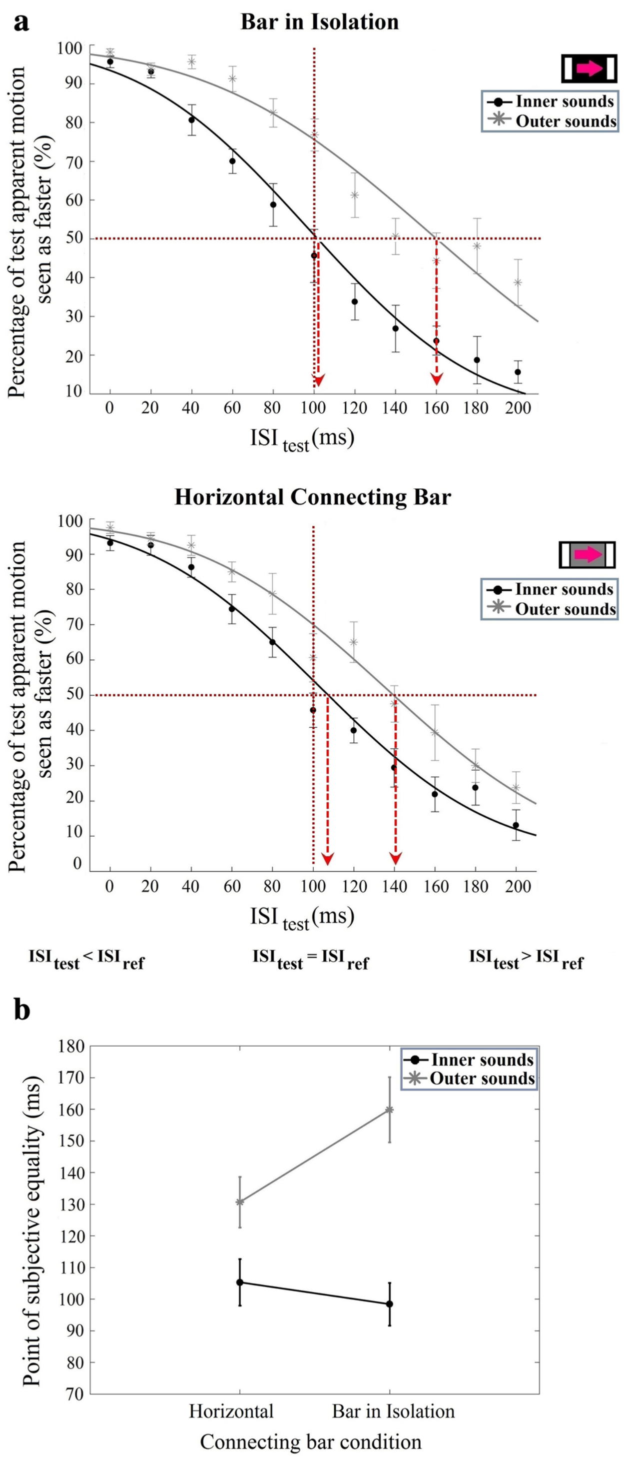


Fig. 8 Results of Experiment 2b ($n=8$). **a** Group-averaged data for different experimental conditions. Each single apparent motion condition is displayed in separate plots, and in each plot, the results for inner and outer sounds are shown. **b** Averaged PSEs of all observers for different experimental conditions. Error bars correspond to ±SEM. Other conventions are the same as those in Fig. 6

values of single-bar conditions in Experiment 1a (Fig. 2 right panel).

General discussion

In the current study, we investigated the relationship between visual spatial grouping principles and audiovisual interactions in time. Accordingly, we examined temporal ventriloquism effects on perceived speed across different visual spatial displays. We used two-frame apparent motion in our experiments and systematically varied the spatial grouping between the flashed bars in each motion frame. We found significant differences between horizontal and vertical connecting bar conditions in motion frames including two or three flashed bars. Overall, the effect of sound timing (inner vs. outer) got smaller when horizontal connecting bars were introduced along the path of apparent motion. For motion frames including two bars, the effect of sound timing did not change significantly across different spatial distance and orientation (i.e., motion direction) conditions. However, for motion frames including three flashed bars, the effects of spatial distance and orientation were found to be significant. When the surrounding bars had different orientation (i.e., motion direction) than the middle one, the effect of sound timing on perceived speed decreased. An increase in spatial distance mostly led to a decrease in the effects of sound timing. Compared to the differences between the connecting bar conditions, these changes in sound timing effects, due to spatial distance and orientation manipulation, were smaller.

Temporal ventriloquism and intramodal grouping

The previous studies have mainly examined temporal ventriloquism by either focusing on the intramodal grouping of audition (i.e., the capturing modality) or manipulating the grouping principles in the temporal domain (e.g., Cook and Van Valkenburg 2009; Keetels et al. 2007). For instance, Keetels et al. (2007) investigated the influence of auditory grouping on temporal ventriloquism. By varying the similarity between the capturing and flanker (background) sounds, they found that temporal ventriloquism effects occurred only when the capturing sounds were perceived as segregated from the flanker sounds. Moreover, the role of temporal proximity has also been examined within the context of apparent motion and it was found to significantly modulate temporal ventriloquism effects (Chen et al. 2011; Roseboom et al. 2013). To complement these previous findings and to better understand the general principles underlying the coordination of intramodal grouping and temporal ventriloquism, we investigated temporal ventriloquism under different spatial grouping conditions in vision. Our results based on the spatial configurations including up to three visual objects showed that visual spatial grouping principles can modulate temporal ventriloquism and audiovisual interactions in time. Thus, these findings suggest that intramodal grouping in visual space can take precedence over temporal ventriloquism. In particular, uniform connectedness had salient and significant effects on temporal ventriloquism in all the apparent motion displays used in the current study. Uniform connectedness has also been found to play a fundamental role in the spatial organization of visual stimuli. Palmer and Rock (1994) argued that uniform connectedness is basic and more prior to other grouping principles such as spatial proximity. When two or more objects are physically connected to each other, they are represented as one group and enter the visual system as a single object. Therefore, the connected objects are primitively grouped, and hence, connectedness can have more dominant perceptual consequences than other spatial grouping strategies. The proposal by Palmer and Rock (1994) has been supported by behavioral studies comparing the effects of different spatial grouping principles on perceptual organization (e.g., Han et al. 1999). By considering the priority given to uniform connectedness over other spatial grouping strategies, it is reasonable to observe that this fundamental spatial grouping can have dominant influences on audiovisual interactions and intersensory pairing in time. In general, manipulations on the spatial location of auditory and visual stimuli did not affect temporal ventriloquism, and the spatial correspondence between the crossmodal stimulations has not been considered a necessary constraint for intersensory pairing in time (Vroomen and Keetels 2006). Moreover, using a single bar in each motion frame, Kafaligonul and Stoner (2010) reported that the effect of spatial displacement between the successive visual flashes of apparent motion did not change the effect of sound timing on motion perception. Our study further suggests that spatial proximity (between each visual object) can only have some significant effects on temporal ventriloquism when three simultaneous objects are presented in each apparent motion frame.

Several studies have also investigated the capacity to bind visual events to a single sound (Olivers et al. 2016; Van der Burg et al. 2013). They reported that audiovisual processing and integration were substantially impaired when more than one visual event was presented with the sound. These findings suggest that audiovisual processing is limited to one visual event and a single auditory event. In the present study, an emerging hypothesis is whether the efficiency of audiovisual interactions in time (i.e., temporal ventriloquism) depends on the number and distinctiveness of visual events in each motion frame. We expect that temporal ventriloquism effects on perceived speed should substantially decrease when the number of visual events in each motion frame and perceived number of moving objects are increased. Even though the observed significant differences between specific conditions can be explained by this prediction, our results do not fully support this view. For instance, in vertical connecting bar conditions, visual stimulus mostly evoked perception of a single and large moving object (in particular when the distance between the connected bars were small), whereas more than one moving object were typically perceived in horizontal connecting bar conditions. In line with the idea that audiovisual processing is limited to one visual event, the effect of temporal ventriloquism on perceived speed was significantly smaller for the horizontal connecting bar conditions than for the vertical connecting bar conditions. On the other hand, this is not the case when we compare single-bar conditions with the corresponding two- or three-bar conditions in which more than one moving object is perceived. Our findings did not indicate a significant decrease in the magnitude of temporal ventriloquism for motion displays including two visual objects (Fig. 2). For displays with three moving objects (Fig. 4), only three conditions of orientation were significantly different than the corresponding single object conditions. It should also be noted certain grouping principles (e.g., similarity/common fate) started to have significant influences on temporal ventriloquism when three apparent motions were presented simultaneously. Furthermore, we found temporal ventriloquism effects on perceived speed in all of the spatial configurations and grouping conditions used. In other words, temporal ventriloquism was not abolished in any of the conditions studied here and the sound timing remained effective on the spatiotemporal processing (i.e., grouping) of apparent motion. Accordingly, our results highlight the presence of temporal ventriloquism in complex spatial configurations. These results are in line with the view provided by behavioral studies using relatively complex auditory and visual stimulation (Vakatis and Spence 2007). For instance, using auditory tone and visual flash sequences, Cook and Van Valkenberg (2009) reported that temporal ventriloquism can be elicited by grouped tone and grouped flash sequences. Of particular relevance to the present study, Kafaligonul and Stoner (2012) found that the time interval demarcated by static sounds (auditory click) can alter sensitivity to apparent motion including multiple visual objects (random dots) in each motion frame (see also Kawachi et al. 2014; Wilbiks and Dyson 2016). As we only used up to three simultaneous moving objects in our experiments, future psychophysical studies are necessary to further understand the relationship between the number of objects in a visual display and audiovisual interactions in time. Characterizing temporal ventriloquism over different number of visual objects and brief sounds awaits further systematic investigation.

Path-guided apparent motion

A faint path (e.g., a gray bar/rectangle), introduced between the successive frames of apparent motion, has been found to alter the spatiotemporal processing of motion and the resulting percept (Shepard and Zare 1983). The temporal and spatial effects induced by a path were shown to correlate with the activity of the primary visual cortex (V1) (Akselrod et al. 2014). This suggests that such intramodal manipulation on apparent motion occurs at early stages of visual processing. Similar experimental designs along the same lines demonstrated the importance of a visual stimulus introduced between successive motion frames as a cue to disambiguate motion direction (Francis and Kim 1999). The analysis and simulations based on dynamic models of vision also illustrated that these effects can be explained by the tuning properties of neurons located at low-level visual areas. In our experiments, we observed that the perceived speed difference between the two sound conditions gets significantly smaller when a horizontal gray bar is introduced between the flashed bars of apparent motion. Thus, our data indicate that the presence of horizontal bar (i.e., path) also affects temporal ventriloquism and audiovisual interactions in time. Based on the findings related to path-guided apparent motion and similar paradigms mentioned above, we expect that the horizontal gray bar may influence and/or engage processing at low-level visual areas. Thus, these intramodal modulations in apparent motion may have some precedence over the information from another modality and over the audiovisual interactions in time. Another important point worth mentioning here is the role of attention in the observed changes due to horizontal connecting bars. In particular, using an experimental design based on spatial ventriloquism, Sanabria et al. (2007) found that spatial cueing can modulate the audiovisual interactions in apparent motion. They suggested that spatial attention reduces intersensory binding by facilitating the segregation of unisensory events. Similarly, compared to the no connecting bar and vertical connecting bar conditions, the brief presentation of horizontal bars (paths) can automatically draw spatial attention to apparent motion, and hence modulate the effects of sound timing and audiovisual interactions in our experiments. Though most of the explanations and computational models of path-guided apparent motion are based on pre-attentive visual processing, we still consider that the presence of a horizontal connecting bars (paths) may act as a spatial cue, and thus engage intramodal grouping of vision and facilitate the segregation of auditory and visual events. Future research examining the role of spatial attention in temporal ventriloquism will be informative in this respect.

Multisensory nature of speed estimation

Speed perception is essential for survival in a dynamic world. By relying on our estimates of speed, we are able to interact with quickly approaching objects and perform the necessary motor actions. The previous motion studies have typically focused on understanding the perception of speed by only using the visual modality. However, the human brain is adapted to operate optimally in natural environments in which behavior is guided by information integrated across multiple sensory modalities. Therefore, such a unisensory approach may not be optimal to tap into natural perceptual mechanisms. Consistent with this perspective, crossmodal interactions have been discovered to have significant influences on motion perception and speed estimation (for reviews see Hidaka et al. 2015; Soto-Faraco et al. 2003). As mentioned in the previous sections, using pairs of auditory clicks and visual flashes, Kafaligonul and Stoner (2010) previously showed that even the timing of brief static sounds can modulate perceived speed. A fundamental issue is whether such effects of sound timing do exist when there is more than one moving object in a visual scene as frequently encountered in our daily lives. Even though the experiments in the current study do not include natural sounds and scenes, our findings contribute to our understanding on this fundamental problem. The results presented here suggest that a static sound can also drive the timing of multiple objects in each motion frame and hence influence the perceived speed of resulting apparent motions. These are in agreement with the previous studies, suggesting that a single auditory event can alter the perception of multiple visual events such as moving objects (e.g., Kawachi et al. 2014). Accordingly, in combination with other studies, our findings on speed estimation here also emphasize the importance of information provided by other modalities for speed estimation in complicated visual scenes. They have implications for designing systems that take the multisensory nature of motion perception and speed estimation into account (Ho and Spence 2008; Horswill and Pooley 2008; Spence and Ho 2008).

Acknowledgements We thank Aaron Clarke and Jennifer Corbett for the discussions on this work and comments on the manuscript. This research was supported by the Scientific and Technological Research Council of Turkey (TUBITAK Grant 113K547).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Akselrod M, Herzog MH, Öğmen H (2014) Tracing path-guided apparent motion in human primary visual cortex V1. Sci Rep 4:6063

Brainard D (1997) The psychophysics toolbox. Spat Vis 10:433–436

Burr D, Banks M, Morrone M (2009) Auditory dominance over vision in the perception of interval duration. Exp Brain Res 198:49–57

Chen L, Shi Z, Müller HJ (2011) Interaction of perceptual grouping and crossmodal temporal capture in tactile apparent motion. PLoS One 6(2):e17130

Cook LA, Van Valkenburg DL (2009) Audio–visual organization and the temporal ventriloquism effect between grouped sequences: evidence that unimodal grouping precedes crossmodal integration. Perception 38:1220–1233

Ekroll V, Faul F, Golz J (2008) Classification of apparent motion percepts based on temporal factors. J Vis 8(4):31. https://doi.org/10.1167/8.4.31

Fendrich R, Corballis PM (2001) The temporal cross-capture of audition and vision. Percept Psychophys 63:719–725

Francis G, Kim H (1999) Motion parallel to line orientation: disambiguation of motion percepts. Perception 28:1243–1255

Freeman E, Driver J (2008) Direction of visual apparent motion driven solely by timing of a static sound. Curr Biol 18:1262–1266

Getzmann S (2007) The effect of brief auditory stimuli on visual apparent motion. Perception 36:1089–1103

Han S, Humphreys GW, Chen L (1999) Uniform connectedness and classical gestalt principles of perceptual grouping. Percept Psychophys 61:661–674

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

Ho C, Spence C (2008) The multisensory driver: implications for ergonomic car interface design. Ashgate Publishing Limited, Aldershot

Horswill MS, Plooy AM (2008) Auditory feedback influences perceived driving speeds. Perception 37:1037–1043

Kafaligonul H, Stoner GR (2010) Auditory modulation of visual apparent motion with short spatial and temporal intervals. J Vis 10(12):31. https://doi.org/10.1167/10.12.31

Kafaligonul H, Stoner GR (2012) Static sound timing alters sensitivity to low-level visual motion. J Vis 12(11):2. https://doi.org/10.1167/12.11.2

Kawachi Y, Grove PM, Sakurai K (2014) A single auditory tone alters the perception of multiple visual events. J Vis 14(8):16. https://doi.org/10.1167/14.8.16

Keetels M, Stekelenburg J, Vroomen J (2007) Auditory grouping occurs prior to intersensory pairing: Evidence from temporal ventriloquism. Exp Brain Res 180:449–456

Kolers PA (1972) Aspects of motion perception. Pergamon Press, Oxford

Kubovy M, Yu M (2012) Multistability, cross-modal binding and the additivity of conjoined grouping principles. Philos Trans R Soc Lond B Biol Sci 367:954–964

Morein-Zamir S, Soto-Faraco S, Kingstone A (2003) Auditory capture of vision: examining temporal ventriloquism. Brain Res Cogn Brain Res 17:154–163

Olivers CNL, Awh E, Van der Burg E (2016) The capacity to detect synchronous audiovisual events is severely limited: evidence from mixture modeling. J Exp Psychol Hum Percept Perform 42:2115–2124

Palmer S, Rock I (1994) Rethinking perceptual organization: the role of uniform connectedness. Psychon Bull Rev 1:29–55

Pelli D (1997) The video toolbox software for visual psychophysics: transforming numbers into movies. Spat Vis 10:437–442

Recanzone GH (2003) Auditory influences on visual temporal rate perception. J Neurophysiol 89:1078–1093

Roseboom W, Kawabe T, Nishida S (2013) Direction of visual apparent motion driven by perceptual organization of cross-modal signals. J Vis 13(1):6. https://doi.org/10.1167/13.1.6

Sanabria D, Soto-Faraco S, Chan JS, Spence C (2004) When does visual perceptual grouping affect multisensory integration? Cogn Affect Behav Neurosci 4:218–229

Sanabria D, Soto-Faraco S, Spence C (2007) Spatial attention and audiovisual interactions in apparent motion. J Exp Psychol Hum Percept Perform 33:927–937

Shepard RN, Zare SL (1983) Path-guided apparent motion. Science 220:632–634

Shi Z, Chen L, Müller HJ (2010) Auditory temporal modulation of the visual Ternus effect: the influence of time interval. Exp Brain Res 203:723–735

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

Spence C (2015) Cross-modal perceptual organization. In: Wagemans J (ed) The Oxford handbook of perceptual organization. Oxford University Press, Oxford, pp 649–664

Spence C, Ho C (2008) Multisensory interface design for drivers: past, present and future. Ergonomics 51:65–70

Takeshima Y, Gyoba J (2013) Complexity of visual stimuli affects visual illusion induced by sound. Vision Res 91:1–7

Ternus J (1926) Experimentelle untersuchungen über phänomenale identität. Psychologische Forschung 7:81–136

Ternus J (1938) The problem of phenomenal identity. In: Ellis WD (ed) A source book of Gestalt psychology. Kegan Paul Trench Trubner & Company, London, pp 149–160

Van der Burg E, Awh E, Olivers CNL (2013) The capacity of audiovisual integration is limited to one item. Psychol Sci 24:345–351

Van der Burg E, Cass J, Alais D (2014) Window of audio–visual simultaneity is unaffected by spatio-temporal visual clutter. Sci Rep 4:5098

Vatakis A, Spence C (2007) Crossmodal binding: evaluating the “unity assumption” using audiovisual speech stimuli. Percept Psychophys 69:744–756

Vroomen J, Keetels MN (2006) The spatial constraint in intersensory pairing: no role in temporal ventriloquism. J Exp Psychol Hum Percept Perform 32:1063–1071

Watanabe K, Shimojo S (2001) When sound affects vision: effects of auditory grouping on visual motion perception. Psychol Sci 12:109–116

Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. Psycholo Bull 88:638–667

Wichmann FA, Hill NJ (2001a) The psychometric function: I. Fitting, sampling and goodness-of-fit. Percept Psychophys 63:1293–1313

Wichmann FA, Hill NJ (2001b) The psychometric function: II. Bootstrap-based confidence intervals and sampling. Percept Psychophys 63:1314–1329

Wilbiks JMP, Dyson BJ (2016) The dynamics and neural correlates of audio–visual integration capacity as determined by temporal unpredictability, proactive interference, and SOA. PLoS One 11(12):e0168304