



## Cortical processes underlying the effects of static sound timing on perceived visual speed



Utku Kaya <sup>a,b</sup>, Hulusi Kafaligonul <sup>a,c,\*</sup>

<sup>a</sup> National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, Turkey

<sup>b</sup> Informatics Institute, Middle East Technical University, Ankara, Turkey

<sup>c</sup> Interdisciplinary Neuroscience Program, Bilkent University, Ankara, Turkey

### ARTICLE INFO

#### Keywords:

Auditory timing  
Perceived speed  
Visual apparent motion  
Audiovisual interactions  
EEG

### ABSTRACT

It is well known that the timing of brief static sounds can alter different aspects of visual motion perception. For instance, previous studies have shown that time intervals demarcated by brief sounds can modulate perceived visual speed such that apparent motions with short auditory time intervals are typically perceived as faster than those with long time intervals. Yet, little is known about the principles and cortical processes underlying such effects of auditory timing. Using a speed judgment paradigm combined with EEG recording, we aimed to identify when and where in the cortex auditory timing takes place for motion processing. Our results indicated significant effects of auditory timing over the medial parieto-occipital and parietal, right centro-parietal, and frontal scalp sites. In addition, these effects were not restricted to a single ERP component and we observed both significant changes in early and late components. Therefore, our findings here suggest that auditory timing may take place at both early and late stages of motion processing and its influences on motion perception may be the outcome of the dynamic interplay between different cortical regions. Together with accumulating evidence, these findings also support the notion that audiovisual integration is a multistage process and it may be achieved through more diversified processes than previously thought.

### 1. Introduction

Motion processing is an important aspect of vision and crucial for survival in a dynamic environment. By relying on our estimates of motion, we are able to interact with quickly approaching objects and perform the necessary motor actions. There is growing interest in crossmodal interactions in visual motion perception (see Soto-Faraco et al., 2003; Hidaka et al., 2015 for reviews). Audiovisual interactions have been heavily studied in this context. Of particular note, static sound timing has been shown to alter the perception of various motion aspects. For instance, the time interval demarcated by brief sounds has been found to alter the perception of apparent motion in distinct features such as category (e.g., element vs. group motion as in Shi et al., 2010; see also Getzmann, 2007), direction (Freeman and Driver, 2008) and speed (Kafaligonul and Stoner, 2010). Also, it has been shown to modulate the sensitivity to motion direction (Kafaligonul and Stoner, 2012). These effects of auditory timing have been mostly explained by the superior temporal resolution of the auditory system (Welch and Warren, 1980; Burr et al., 2009) and by a phenomenon called temporal ventriloquism, in

which brief sounds drive the perceived timing of visual events (Fendrich and Corballis, 2001; Morein-Zamir et al., 2003; Recanzone, 2003). More specifically, these time interval effects have been described as brief sounds capturing the timing of each apparent motion frame and hence leading to perceptual changes in different motion features.

In particular, temporal ventriloquism effects on perceived speed have been found to be robust. In these studies, two-frame apparent motion and two auditory clicks temporally centered with the apparent motion were typically used. When the time interval between auditory clicks was smaller than the time interval between apparent motion frames, subjects consistently judged the apparent motion to be moving faster than the one presented with clicks of longer time intervals (Kafaligonul and Stoner, 2010). As in other motion studies mentioned above, these changes have been mainly explained by describing that brief sounds capture each motion frame in time and drive the timing of these visual events (or the time interval between these events). Therefore, the shortening and lengthening in the time interval between motion frames have been considered to result in faster and slower motion percepts, respectively. The position of the sound source had little or no influence on these effects

\* Corresponding author. Aysel Sabuncu Brain Research Center, Bilkent University, Ankara, 06800, Turkey.

E-mail address: [hulusi@bilkent.edu.tr](mailto:hulusi@bilkent.edu.tr) (H. Kafaligonul).

(Vroomen and Keetels, 2006; Kafaligonul and Stoner, 2010; Ogulmus et al., 2018). Auditory time intervals can also affect the perceived speed of apparent motion displays that include more than one object in each motion frame having different spatial configurations (Ogulmus et al., 2018). Overall, these behavioral studies provide strong evidence that auditory time intervals play an important role in shaping perceived visual speed. However, both cortical and subcortical processes underlying such temporal effects are quite poorly understood.

Kafaligonul and Stoner (2010) discovered that auditory time intervals can modulate the perceived speed of apparent motions with short spatial and temporal offsets. Such apparent motion stimuli are thought to mostly engage low-level motion areas such as middle temporal area (area MT). Therefore, these behavioral findings suggest that these audiovisual interactions in time are not restricted to high-level motion areas and auditory timing may also be used at early stages of motion processing. On the other hand, Kafaligonul et al. (2018) have recently found no obvious correlate of temporal ventriloquism effects in area MT of awake fixating macaques, although they showed that auditory clicks can modulate area MT activity. In another recent study, Kaya et al. (2017) examined the effects of adaptation to specific auditory time-intervals on the evoked activity elicited by visual apparent motion. Their results indicated significant aftereffects on the event-related potentials (ERPs) over parietal electrodes. The amplitudes of early components (50–80 ms and 140–180 ms time ranges) were significantly changed through adaptation to auditory time-intervals and these changes were also extended over occipital scalp regions. The aftereffects on the later (>300 ms) ERP components were more salient and mostly centered over right centro-parietal electrodes. An emerging hypothesis based on these findings is that auditory timing may be progressively used for visual motion processing over these areas, starting from parieto-occipital up to centro-parietal sites. It should also be noted that participants performed a simple motion direction discrimination task just to draw their attention to visual motion in these EEG recordings. In addition, the behavioral performance was high for all conditions and there was no robust difference between auditory adaptation conditions. Therefore, while such experimental design was highly informative about the nature of sub-second time interval adaptation and the aftereffects on the evoked activity, the findings did not provide a direct relationship between changes in speed perception and distinct ERP components.

Accumulating evidence also indicates that multisensory integration is a multistage process and involves cortical networks operating at distinct stages of perceptual processing (Talsma et al., 2010; Murray et al., 2016). It has been suggested that these different processes may be adaptively recruited based on the nature of sensory stimulation and specific task demands (Senkowski et al., 2007; van Atteveldt et al., 2014). For instance, Cecere et al. (2017) have recently shown that even a change in the relative timing (i.e., temporal order) of auditory and visual stimuli can engage different audiovisual processes operating over distinct ERP components. In these experiments (see also Cecere et al., 2016), they used an audiovisual simultaneity paradigm with a single brief sound and a single visual flash. Based on the leading modality in time (auditory-leading vs. visual-leading stimulus pairs), they found that the organization of audiovisual interactions differed in terms of scalp activity patterns within distinct time-ranges. These findings were interpreted as the involvement of different networks and mechanisms based on the sensory system engaged first. In other words, different mechanisms of crossmodal interaction may come into play depending on which signal is cueing the follow-up processing. The leading auditory stimulus may alert the visual system of imminent input (i.e., bottom-up attention), whereas the visual stimulus may have a predictive value for a forthcoming auditory input (Thorne and Debener, 2014). Although such interpretation may not be directly applied to relatively complex stimulation in the paradigms demonstrating temporal ventriloquism effects on visual apparent motion, it is possible that auditory clicks may also engage mechanisms based on predictions. From an ecological perspective, moving objects in the environment mostly produce concurrent sounds.

Similar to other visual and auditory attributes (Spence, 2011; Orchard-Mills et al., 2016), correspondences between different aspects of motion and auditory timing may exist. For instance, hearing clicks with a high flutter (short time interval) or low flutter (long time interval) rates may have predictive power for faster or slower moving objects, respectively. Such predictive power may engage high-level processing for motion and speed estimation. Although these previous studies overall suggest the involvement of different cortical processes in the effects of concurrent stimulation on perceived speed, the basic properties (time frame and scalp sites) of such cortical processes remain unknown.

In the current study, we focused on understanding when and where in the cortex auditory timing (and time intervals) take place for perceived visual speed and hence attempted to reveal the cortical processes underlying the multisensory nature of speed estimation. We acquired both behavioral data and EEG (Electroencephalography) activity in tandem while participants compared the speed of two consecutive apparent motions presented with different auditory time intervals (i.e., two-interval forced-choice task on visual speed). Building from previous work mentioned above, we expected to find significant effects of auditory timing at both early and later stages of cortical processing. We further predicted that different temporal parameters (i.e., experimental factors) defined according to the relative timing of each auditory click and each motion frame, would take place over distinct ERP components and electrode sites. To systematically manipulate the audiovisual interactions specific to auditory timing, we defined two auditory time intervals (short vs. long) relative to the one between apparent motion frames. Additionally, we varied the level of temporal disparity between these auditory time intervals. As the temporal disparity level was increased, the difference between the two time intervals became obvious and thus the predictive power of clicks for speed estimation became high. Therefore, we expected that the temporal disparity level would engage auditory modulations at later stages of motion processing. On the other hand, since an increase in disparity level also led to an increase in the time interval between each click and apparent motion frame (i.e., a decrease in temporal proximity between each click and each visual flash), such manipulation was expected to decrease audiovisual interactions at early stages of sensory processing. As also mentioned above, compared to vision, audition has been found to have a dominant role in temporal processing and tasks dependent on timing (Chen and Vroomen, 2013). Although such priority of audition has been well studied at the perceptual level, our understanding of underlying neural mechanisms (in particular within the sub-second range) is limited. The experimental paradigm studied here provides a great example of such auditory dominance in time. Therefore, our systematic manipulations and findings here provide additional insight into the mechanisms underlying crossmodal temporal processing within the sub-second range.

## 2. Methods

### 2.1. Participants

Twenty-three healthy volunteers participated in this study. The data of 4 participants were excluded from analysis because either their performances did not meet our criteria in speed judgments (see *Task and procedure*) or they had excessive EEG artifacts. Thus, the data from 19 participants were included in the analysis (7 females, 18 right-handed, age range 19–32 years). All participants had normal or corrected-to-normal visual acuity and normal hearing. None of them had a history of neurological disorders. They provided written informed consent, and all procedures were in accordance with the Declaration of Helsinki (World Medical Association, 2013) and approved by the local ethics committee at the Faculty of Medicine, Ankara University.

### 2.2. Apparatus and stimuli

We used MATLAB version 7.12 (The MathWorks, Natick, MA) with

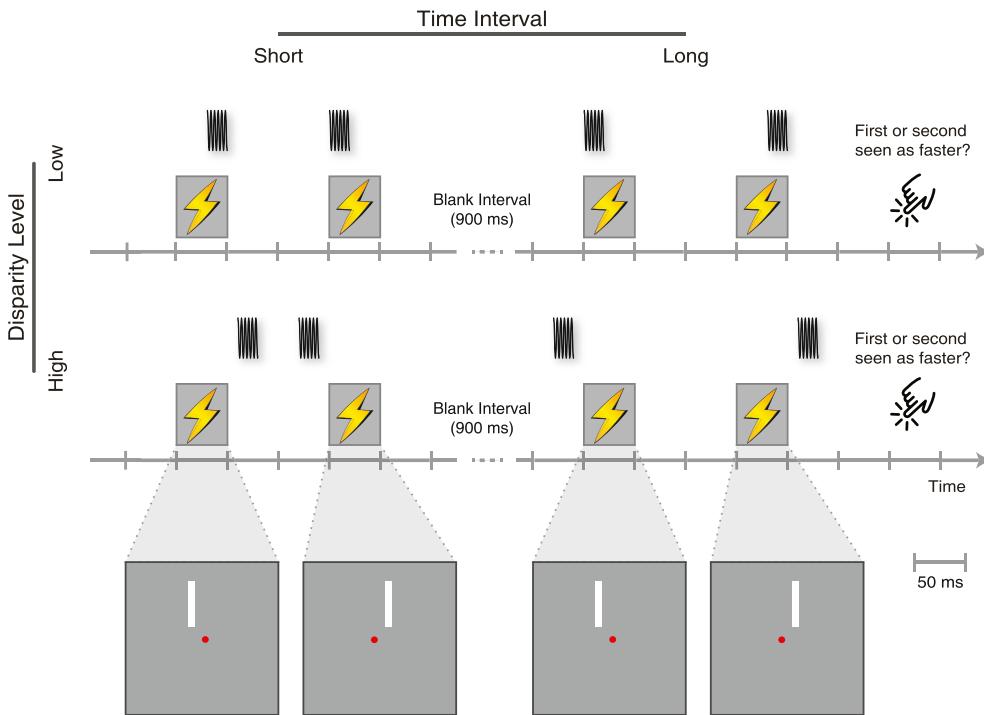
the Psychtoolbox 3.0 for stimulus presentation and data acquisition (Brainard, 1997; Pelli, 1997). Visual stimuli were presented on a 21-inch CRT monitor ( $1280 \times 1024$  pixel resolution and 100 Hz refresh rate) at a viewing distance of 57 cm. Luminance values were measured by a SpectroCAL (Cambridge Research Systems, Rochester, Kent, UK) photometer. A gamma-corrected lookup table (LUT) was used so that luminance was a linear function of the digital representation of the image. Sounds were introduced via insert earphones (EARTone 3A, Etymotic Research, Village, IL) and amplitudes were measured by a sound-level meter (SL-4010, Lutron Electronics, Taipei, TW). The 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). All experimental sessions were performed in a silent, dimly lit room.

A small red circle (0.3 deg diameter) at the center of the display served as a fixation point. We used two-frame apparent motion for visual stimulation (Fig. 1). In each motion frame, a single bar ( $0.7 \times 2.7$  deg with a luminance of  $97 \text{ cd/m}^2$ ) was “flashed” for 50 ms on a gray background ( $20 \text{ cd/m}^2$ ). The bar locations were adjusted in the upper visual field such that the center of the apparent motion was located 3.4 deg above the fixation point. The spatial displacement (i.e., center-to-center horizontal separation) and the inter-stimulus interval (visual ISI<sub>v</sub>) between each flashed bar were 1.3 deg and 100 ms, respectively. The direction of motion was either rightward or leftward. Auditory stimuli were a pair of static clicks. Each click had a 20 ms duration, was comprised of a rectangular windowed 480 Hz sine-wave carrier and was sampled at 44.1 kHz with 8-bit quantization. The clicks were introduced binaurally at 75 dB sound pressure level (SPL).

### 2.3. Task and procedure

During each trial, the two-frame apparent motion stimulus was presented twice and the blank interval between each presentation (i.e., the interval between the offset of the first and the onset of the second presentation) was 900 ms (Fig. 1). In terms of visual stimulation, each consecutive presentation was exactly the same, but the timing of the auditory stimulation differed. A pair of clicks were centered temporally

with each presentation of apparent motion. For one of the successive apparent motions, the time interval (auditory ISI<sub>a</sub>) between them was either shorter than or equal to the time interval between each motion frame (*short* condition). For the other one, auditory time interval was longer than the time interval between each motion frame (*long* condition). The temporal order of *short* and *long* conditions was randomized across trials. In addition to these basic auditory time interval conditions defined relative to the visual time interval, we also systematically varied the level of temporal disparity between them. In the *low* disparity condition, the difference between two auditory time intervals in the same trial was low (*short* ISI<sub>a</sub>: 100 ms, *long* ISI<sub>a</sub>: 160 ms). As shown in Fig. 1, each click was overlapping with the presentation of each bar and either the onset or the offset of each click matched with the onset or offset of each bar. On the other hand, for the *high* level of disparity, the difference between two auditory time intervals was high (*short* ISI<sub>a</sub>: 40 ms, *long* ISI<sub>a</sub>: 220 ms). Each click was not overlapping with the presentation of each bar in time. For both *short* and *long* conditions of *high* disparity level, the ISI between each click and each bar (i.e., the first click-the first bar or the second click-the second bar) was 10 ms. In addition to these audiovisual (AV) conditions (2 auditory time intervals x 2 level of disparity), we also included 4 auditory-only (A) and 1 visual-only (V) conditions. Except presenting either auditory or visual stimuli, we used the same stimulus parameters of the AV conditions in these unimodal conditions. Since two successive apparent motions were exactly the same during each trial, the number of trials for the visual-only condition was half of the other conditions so that the number of stimulus presentations was the same for all conditions. All of these conditions were pseudorandomly intermixed and presented 12 times during each experimental session. At the end of each trial, observers indicated, by pressing one of the two keys, which of the two consecutive apparent motions appeared to move faster (two-interval forced-choice). After the keyboard press and a variable inter-trial interval (0.5–1.5 s), during which only the fixation point was present, the next trial started. Observers were told that the visual stimuli would be accompanied by clicks but to base their responses solely on the visual stimuli. They were asked to only fixate and not to respond when there were not any moving stimuli during a trial (i.e., auditory-only conditions). In other words, as in bimodal (AV) conditions, observers passively listened to clicks in these auditory-only trials. As will be detailed below



**Fig. 1. Experimental design and timeline of each trial.** During each trial, a two-frame apparent motion was presented to the observer twice with a temporal delay of 900 ms. In terms of visual stimulation, these consecutive apparent motions were exactly the same. However, the time interval (ISI<sub>a</sub>) between static clicks was either shorter or equal to (short time interval), or longer (long time interval) than the ISI<sub>v</sub> between the apparent motion frames. At the end of each trial, observers were asked to report which apparent motion moved faster. Temporal order of time interval conditions was randomized and the fixation circle was present throughout the trial. The difference between the two time interval conditions was small and large for low and high temporal disparity levels, respectively. These disparity level conditions were introduced in separate trials. The apparent motion frames and auditory clicks are represented by flash and sinusoidal-waveform icons, respectively. Relative durations of visual and auditory events are indicated by the thickness of these icons (height of these icons distinguishes stimulus modality and is otherwise irrelevant).

(see *Standard ERP analyses*), our ERP analyses were based on testing the additive model [(AV-A) vs. V or AV vs. (A + V)]. By instructing observers not to perform any task and passively listen to clicks in the auditory trials, major confounding factors were circumvented in these analyses. For example, if observers performed a task in the auditory-only trials, the motor response in the AV trials would be subtracted by the motor response in the (A) trials and the difference (AV-A) ERP would contain no motor response. It would be unfair to compare these difference ERPs with that of visual-only (V) which contains a motor response. A similar problem occurs when comparing the summed ERPs [i.e., (A + V) summed ERP which contains the summation of two motor responses] with the corresponding AV (contains only one motor response) ERPs.

Against our instructions, observers could have conceivably ignored apparent motions and relied only on auditory time intervals for their speed judgments in the audiovisual conditions. There are specific reasons for believing that this is unlikely. First, the participants were trained in speed judgments and completed practice sessions including visual-only stimulation without any sound or feedback. After these practice sessions, they started the main EEG experimental sessions lasting at most 20 min in total (see *below*). During this time period, it seems unlikely that they could learn how to use auditory timing and change their strategy for speed judgments on apparent motion completely. Second, in a previous behavioral study, a very similar procedure was also used to assess the temporal ventriloquism effects on perceived visual speed (Ogulmus et al., 2018). Through an additional control experiment of that study (Exp. 2b), it was confirmed that the measurements based on the comparison of two apparent motions with different auditory time intervals (AV short vs. AV long) in a single trial are reflecting changes in perceived visual speed. In the control experiment, instead of providing two auditory time interval conditions (AV short vs. AV long) consecutively in each trial, participants compared the speed of each time interval condition (i.e., AV short or AV long) to a common visual reference (i.e., apparent motion without clicks). The visual ISI of the common reference was changed across different trials. Accordingly, we expect such design to be more resistant to any confounding factor due to response/decisional biases on auditory time intervals. Based on the comparison of each auditory time interval condition with the common reference, psychometric curves were fitted to estimate the point of subjective equality (PSE) and perceived visual speed. The changes in PSE values were in the same direction and in agreement with the changes observed in the design based on the comparison of two sequentially presented apparent motions with different auditory time intervals. In any case, to make sure that observers performed the task according to our instructions in the main EEG experimental sessions, we also included 12 catch trials in these sessions. In each of these catch trials, we presented both modalities as in the audiovisual conditions. However, the auditory time interval for each consecutive presentation was exactly the same ( $ISI_a = 130$  ms). The visual time intervals ( $ISI_v$ ) used for each successive apparent motion were 40 and 160 ms. The order of these visual time intervals was randomized across trials. The auditory time interval was longer and shorter than 40 and 160 ms visual time intervals, respectively. Even though it was expected to have temporal ventriloquism and this effect of clicks was expected to reduce the perceived speed difference between two apparent motions, the difference between the two visual time intervals was high such that the apparent motion with 40 ms  $ISI_v$  was typically perceived much faster than the other one. Therefore, an observer who performed the speed discrimination task according to the instructions should typically have reported the apparent motion with 40 ms  $ISI_v$  faster than the other one. Otherwise, an observer whose criteria was only based on auditory time intervals should not have reported a significant difference between the speeds of two apparent motions.

Each observer completed four experimental sessions for a total of 48 presentations per condition. They were encouraged to have a short break (1–2 min) between the sessions to maintain high concentration and prevent fatigue. Prior to these experimental sessions, each participant was shown examples of apparent motion stimuli followed by practice

(i.e., training) sessions. In the practice sessions, we only used the visual-only condition of the main experiment. One of the apparent motions had a fixed  $ISI_v$  of 100 ms, and the other had a variable  $ISI_v$  between 40 and 190 ms for each trial. The order of these apparent motions was randomized from trial to trial and observers compared their speed at the end of each trial. Participants who had low-performance values in the practice sessions were not included in the main experimental sessions. Moreover, participants, who did not report the apparent motion with 40 ms  $ISI_v$  moved as faster in the majority of catch trials, were also excluded from further data analysis.

#### 2.4. EEG data acquisition and preprocessing

Procedure for EEG data acquisition and preprocessing steps were similar to those in our previous study (Kaya et al., 2017). EEG signals were recorded with a 64-channel MR-compatible system (Brain Products, GmbH, Gilching, Germany), using sintered Ag/AgCl passive electrodes mounted on an elastic cap. The placement of electrodes on the cap was according to the international 10/20 system. Two of the scalp electrodes, FCz and AFz, were used as the reference and ground electrodes, respectively. A conductive paste (ABRALYT 2000, FMS, Herrsching-Breitbrunn, Germany) was applied to reduce impedances in each electrode. Throughout experimental sessions, electrode impedances were kept below 20 k $\Omega$  (typically below 10 k $\Omega$ ) and monitored for reliable recording. EEG data were acquired at a sampling rate of 5 kHz and filtered online by using a band-pass filter (0.016–250 Hz). BrainVision Recorder Software (Brain Products, GmbH) was used to store stimulus markers and EEG data on a secure hard disk for further analyses.

EEG data analyses were carried out offline with BrainVision Analyzer 2.0 (Brain Products, GmbH), the Fieldtrip toolbox (Oostenveld et al., 2011) and our own custom MATLAB scripts (The MathWorks). In preprocessing, EEG signals were first down-sampled to 500 Hz, and the cardioballistic artifacts were removed using the signal from the ECG channel (Allen et al., 1998). Then, the data were filtered through a zero-phase shift Butterworth high-pass filter (0.5 Hz, 24 dB/octave) and a 50-Hz notch filter (50 Hz  $\pm$  2.5 Hz, 16th order) to remove the noise from the power line. After the initial filtering procedure, data were segmented into epochs from -400 ms (before the onset of each apparent motion) to 1000 ms (after the onset of each apparent motion). Each trial (i.e., presentation/epoch) was screened automatically by artifact rejection criteria as well as manually by eye. In the automatic artifact rejection, any trial with oscillations over 50  $\mu$ V/ms, a voltage change of more than 200  $\mu$ V in a 200 ms time window or less than 0.5  $\mu$ V in a 100 ms time window was rejected. Then, an infomax independent component analysis was applied to these epochs to remove common EEG artifacts such as eye blinks. Bad channels were corrected using topographic spline interpolation (Perrin et al., 1989). Trials with artifacts were rejected from further analyses. After applying standard preprocessing procedures, on average 2.37% of trials were rejected per condition. The percentage values for each condition are presented in [Supplementary Table S1](#).

#### 2.5. Standard ERP analyses

EEG signals from each specific electrode location were averaged across trials to compute event-related potentials (ERPs) time-locked to the onset of visual apparent motion (and to the corresponding time point in auditory-only conditions). These averaged ERPs were filtered with a low-pass filter to further smooth the signals (40 Hz cut-off frequency). Baseline correction was applied according to the activity in the -145 to -45 ms preceding the onset of each apparent motion. Within this time range, there was no stimulation (i.e., no auditory clicks or apparent motion frames) for all the experimental conditions.

In the specific audiovisual paradigm studied here, observers performed a visual task (i.e., speed judgment) and listened to sounds passively. Visual apparent motion and auditory clicks acted as primary and secondary task-irrelevant stimuli, respectively. Such experimental

design implies that the information provided by audition (secondary modality) interferes and interacts with the motion/speed estimation primarily carried out by the visual system. Therefore, in our ERP analyses, we relied on the application of an additive model to detect nonlinear neural response interactions and to reveal modulations of these nonlinear components by auditory timing (Stevenson et al., 2014 for a review and comparison of models). In these ERP analyses (e.g., Mishra et al., 2007; Cappe et al., 2010; Naeu et al., 2011), the ERPs in response to the bimodal (AV) conditions are compared with the synthetic summed ERPs in response to the corresponding auditory-only (A) and visual-only (V) conditions. This is also equivalent to comparing synthetic difference (AV-A) ERP with that of visual-only. Accordingly, to identify significant changes specific to auditory timing and time intervals, we subtracted the averaged ERPs of the auditory-only (A) conditions from those of the corresponding audiovisual (AV) conditions.

## 2.6. Statistical analyses

Our statistical analyses were oriented to reveal the main effects of each temporal factor (Time Interval: short vs. long, Disparity Level: low vs. high) and two-way interaction. We relied on paired t-tests and ANOVAs on the difference (AV-A) ERPs. It is important to note that a paired t-test or an ANOVA test on the (AV-A) difference ERPs leads to the same statistical results as the one on the [AV- (A + V)] difference ERPs since exactly the same visual-only (V) data point is subtracted from the four (AV-A) conditions in the later one. Moreover, these tests were designed to reveal significant changes specific to auditory timing in these bimodal difference (AV-A) ERPs rather than comparing a specific difference ERP to that of visual-only (i.e., a baseline). Any confounding factor that existed in all the bimodal conditions (i.e., in all the difference ERPs) and did not change with auditory timing were not reported as significant. Hence, they are expected to be resistant to any confounding factor such as common anticipatory processes that might lead to spurious early interactions (Teder-Sälejärvi et al., 2002; Besle et al., 2004).

Cluster-based permutation test was carried out at the first stage of our statistical analyses. In this analysis, a data-driven non-parametric framework is used to solve the problem of multiple statistical comparisons and to cluster selected samples objectively (Maris and Oostenveld, 2007; Groppe et al., 2011). In brief, two conditions were compared at each electrode location and each time point via paired samples t-test ( $\alpha = 0.05$ ). All significant samples (electrode, time point) were clustered on the basis of spatial and temporal proximity and the t-values within a cluster were summed to have cluster-level statistics. We required at least three neighboring electrodes to form a cluster. To generate the null hypothesis distribution of the cluster-level statistics, this procedure was repeated using 10,000 random permutations of the original data and with the help of the Monte Carlo method. A cluster in the real data was considered to be significant when it fell in the highest or the lowest 2.5<sup>th</sup> percentile of the generated distribution (corresponding to the significance level of a two-tailed test). Since this test can only compare two conditions at a time, we used derived waveforms by combining or subtracting the difference (AV-A) ERPs across different conditions. For the main effect of time interval, the difference ERPs of disparity levels were combined (i.e., averaged) and then, these two combined waveforms were provided as input to the cluster-based permutation test (combined short<sub>AV-A</sub> vs. combined long<sub>AV-A</sub>). The difference ERPs of time intervals were combined at each disparity level and compared (combined low<sub>AV-A</sub> vs. combined high<sub>AV-A</sub>) for the effect of disparity level. The differential effects of time interval at each disparity level (i.e., two-way interaction) were illustrated by subtracting the difference ERPs of time intervals (short<sub>AV-A</sub> – long<sub>AV-A</sub>) at each disparity level and a cluster-based permutation test was used to compare these subtracted waveforms (subtracted low<sub>AV-A</sub> vs. subtracted high<sub>AV-A</sub>). In our previous study, we used a two-frame apparent motion with highly similar parameters (Kaya et al., 2017). More important, the timeline of apparent motion was exactly the same as the one used here. The results indicated that two-frame apparent

motion elicits strong P1, N1 components, later positivity around 300 ms and beyond 400 ms. We focused on these major components reported by prior research. The cluster-based permutation tests and comparisons were performed in 4 different time-ranges (0–140 ms, 140–240 ms, 240–380 ms, 380–550 ms post-stimulus onset) which were arranged to cover conveniently each of these major components.

Based on the outcome of cluster-based permutation tests, we identified significant effects of auditory timing and spatiotemporal clusters associated with these effects. To display evoked brain activity time-courses for illustrative purposes, we also identified time windows and cluster of electrodes (i.e., exemplar sites) over which these spatiotemporal clusters were mainly located. For the identified time windows and exemplar sites, we computed the averaged ERP amplitudes (i.e., AV-A amplitudes) of each participant and performed two-way repeated measures ANOVA, with time interval and disparity level as factors, on these averaged values. To understand the nature of a significant two-way interaction, post-hoc paired t-tests were applied. Moreover, we compared the averaged ERP amplitude of each condition (i.e., AV-A of each timing/time interval condition) with that of visual-only (V) by using paired t-tests. Any significant deviation from the visual-only condition indicated a super-additive [AV > (A + V)] or a sub-additive interaction [AV < (A + V)].

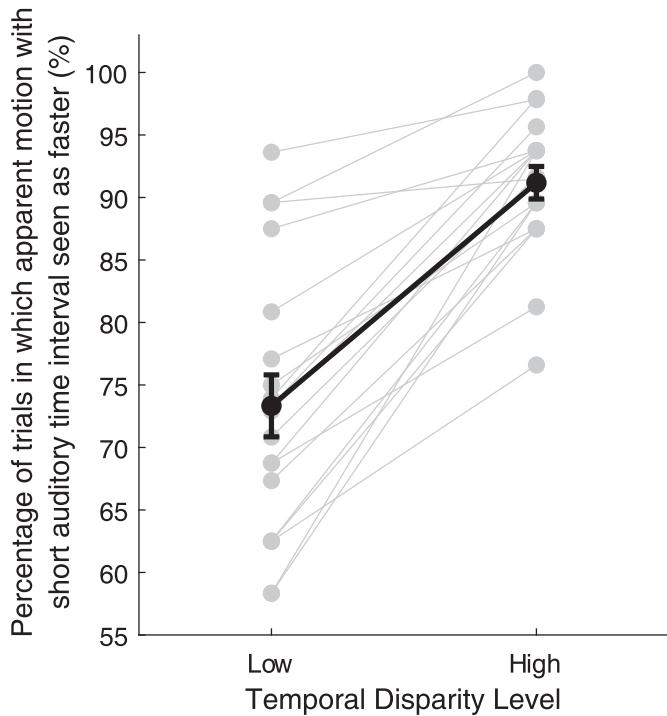
## 2.7. Source localization analyses

Similar to previous studies (e.g., Urgen et al., 2018), the derived (e.g., combined or subtracted) waveforms were used to locate neural generators for scalp topographies in three-dimensional cortical space. The Standardized Low-Resolution Brain Electromagnetic Tomography (sLORETA) technique was employed (Pascual-Marqui, 2002). sLORETA divides intra-cerebral volume into 6239 voxels with 5 mm spatial resolution. The standardized current density at each voxel is computed within a realistic head model using the Montreal Neurological Institute (MNI152) template (Mazziotta et al., 2001). This computation is based on a linear weighted sum of the scalp electric potentials. sLORETA has been proven to achieve reliable localization of possible sources and the validity of these sources was confirmed by other techniques such as functional magnetic resonance imaging (Sekihara et al., 2005; Hoffmann et al., 2014). Within the identified time windows, the source estimations were performed for each participant and each derived waveform. sLORETA images were compared across experimental conditions (Time Interval: combined short<sub>AV-A</sub> vs. combined long<sub>AV-A</sub>; Disparity Level: combined low<sub>AV-A</sub> vs. combined high<sub>AV-A</sub>; Interaction: subtracted low<sub>AV-A</sub> vs. subtracted high<sub>AV-A</sub>), using built-in voxel-wise randomization tests with 5000 permutations based on statistical non-parametric mapping (Nichols and Holmes, 2002).

## 3. Results

### 3.1. Behavioral results

The trials excluded during the EEG preprocessing stage were not used for the analysis of the behavioral data, and the average percentage values from each subject and group-averaged data are shown in Fig. 2. In agreement with the previous findings (Kafaligonul and Stoner, 2010; Ogulmus et al., 2018), the apparent motion with short auditory time interval was perceived to move faster than the one with a long time interval in more than 60% of the trials, although these apparent motions were identical in terms of visual stimulation. For both disparity levels, the percentage values were significantly higher than the 50% chance level (two-tailed t-test using Bonferroni adjusted  $\alpha = 0.05/2$ ; low disparity:  $t_{18} = 9.413, p < 0.001$ ; high disparity:  $t_{18} = 31.613, p < 0.001$ ). Compared to the low disparity condition, the percentage value for the high disparity level was significantly higher ( $t_{18} = 8.468, p < 0.001$ ). This indicates a significant amplification and modulation of the perceived speed difference between short and long time interval conditions by



**Fig. 2. Behavioral data ( $n = 19$ ).** The percentage of apparent motion with short auditory time interval seen as faster is displayed as a function of temporal disparity level. The data from individual participants are shown by gray, and the group-averaged data are displayed by black symbols. Error bars correspond to  $\pm$  SEM.

temporal disparity, and hence suggests a significant interaction between the two factors. With regards to the visual-only condition, in which there were no auditory clicks and physically identical apparent motions were compared, participants perceived the first and second apparent motions in time as almost having the same speed. The percentage of trials in which the first apparent motion seen as faster was close to 50% chance ( $M = 47.77\%$ ,  $SEM = 5.33\%$ ) and it was not significantly different than this level ( $t_{18} = 0.419$ ,  $p = 0.681$ ). An additional analysis on catch-trials, in which the auditory time intervals were the same but the time interval between the apparent-motion frames ( $ISI_v$ ) differed, was carried out. The percentage of catch-trials in which the apparent motion with short visual time interval seen as faster was significantly higher than the 50% chance ( $M = 70.75\%$ ,  $SEM = 2.49\%$ ,  $t_{18} = 8.349$ ,  $p < 0.001$ ). Overall, this confirms that participants performed speed discrimination according to the instructions and suggests that any decisional bias on auditory time intervals (e.g., using only auditory time interval and completely ignoring visual apparent motions during speed discrimination) was limited.

### 3.2. Effects of auditory timing on the ERPs: time-course, scalp topographies, and source estimations

A cluster-based permutation test in the N1 component range revealed a significant effect of time interval (short vs. long). Within this time range, the average activities for short auditory time intervals were more negative (larger N1 amplitude) compared to those of long time intervals (Fig. 3). The significant difference was present at the right hemisphere and mainly clustered over centro-parietal scalp sites in the 150–200 ms time range (cluster-level  $t_{sum} = -650.559$ ,  $p = 0.017$ ). The time interval effect also spread over parietal, central, and temporo-parietal electrodes (Fig. 3A). An additional sLORETA analysis comparing the short and long time interval conditions suggested that this effect was associated with changes in the right parietal (inferior parietal lobule, postcentral gyrus), temporal (superior and transverse temporal gyri), and insular cortices. Compared to the time interval factor, the significant effects of disparity

level (low vs. high) occurred in a later component (~300 ms, cluster-level  $t_{sum} = 963.389$ ,  $p = 0.024$ ). In the 260–310 ms time range, the effect was most pronounced over right frontal and fronto-central electrodes and the averaged activities for low disparity conditions were higher than those of high disparity conditions (Fig. 3A and B). Although there was a similar modulation over some of the left frontal electrodes, the main effect of disparity level was stronger and dominant at the right hemisphere. The significant spatiotemporal cluster associated with this effect even included some of the right fronto-temporal and temporal electrodes. The supplementary source localization analysis pointed to changes in the activations of right superior frontal gyrus.

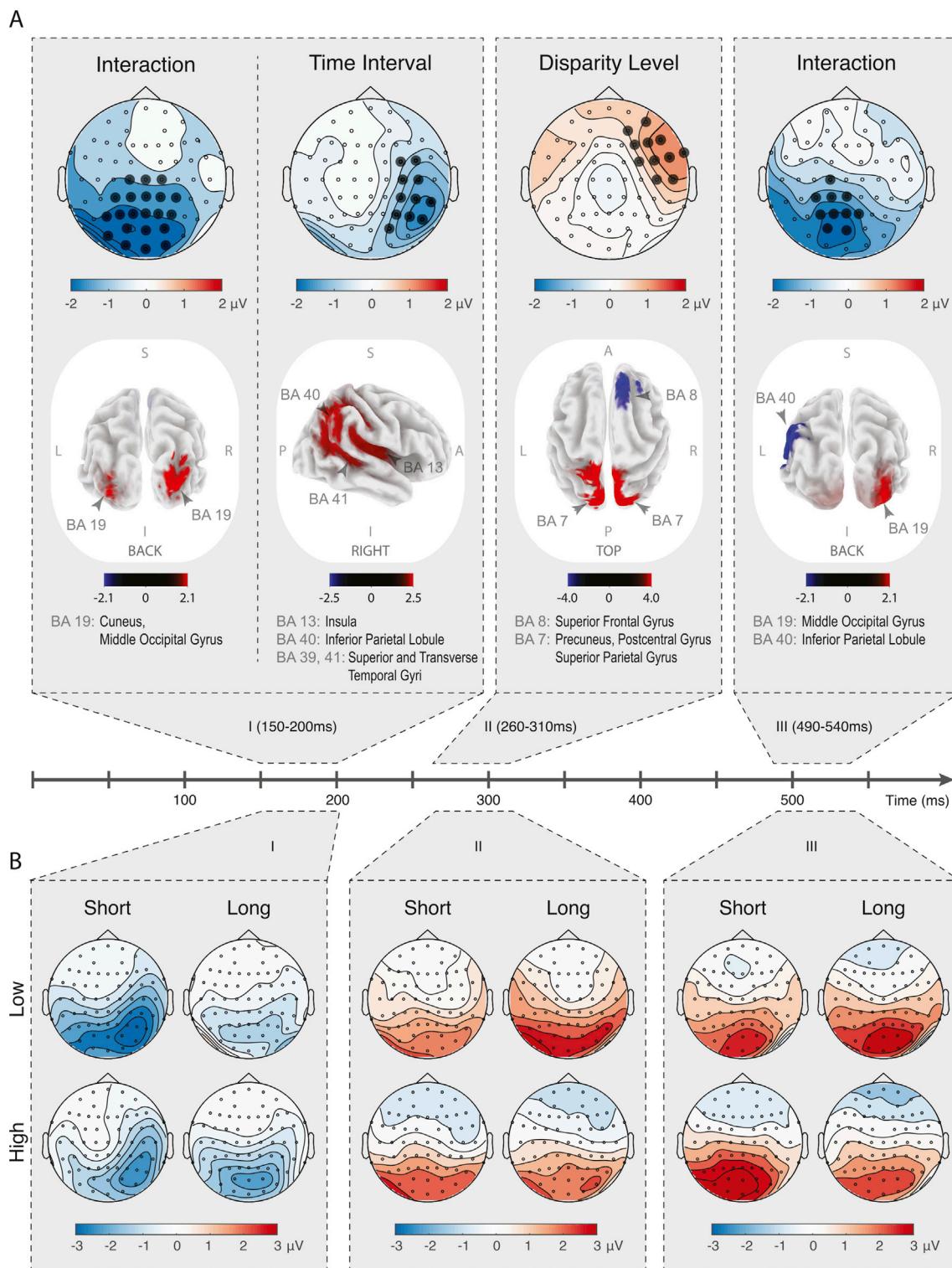
The differential effects of the time interval for each disparity level were revealed through the interaction between the two factors. There were early (150–200 ms time range, cluster-level  $t_{sum} = -1066.458$ ,  $p = 0.010$ ) and late (490–540 ms time range, cluster-level  $t_{sum} = -674.945$ ,  $p = 0.023$ ) significant two-way interactions. For both time windows, the effect of time interval was in the opposite direction for each disparity level (Fig. 3A and B). In terms of scalp topography, the significant two-way interactions were mostly clustered over medial parietal electrodes and also extended over occipital and central regions. The additional analyses suggested that early (150–200 ms) two-way interaction was due to the sources localized in the middle occipital gyrus and cuneus. Moreover, they revealed that the modulations in the middle occipital gyrus were also associated with the later (490–540 ms) two-way interaction. The cluster-based permutation tests did not reveal any other significant spatiotemporal cluster.

### 3.3. Effects of auditory timing on the ERPs: averaged ERP amplitudes from exemplar sites

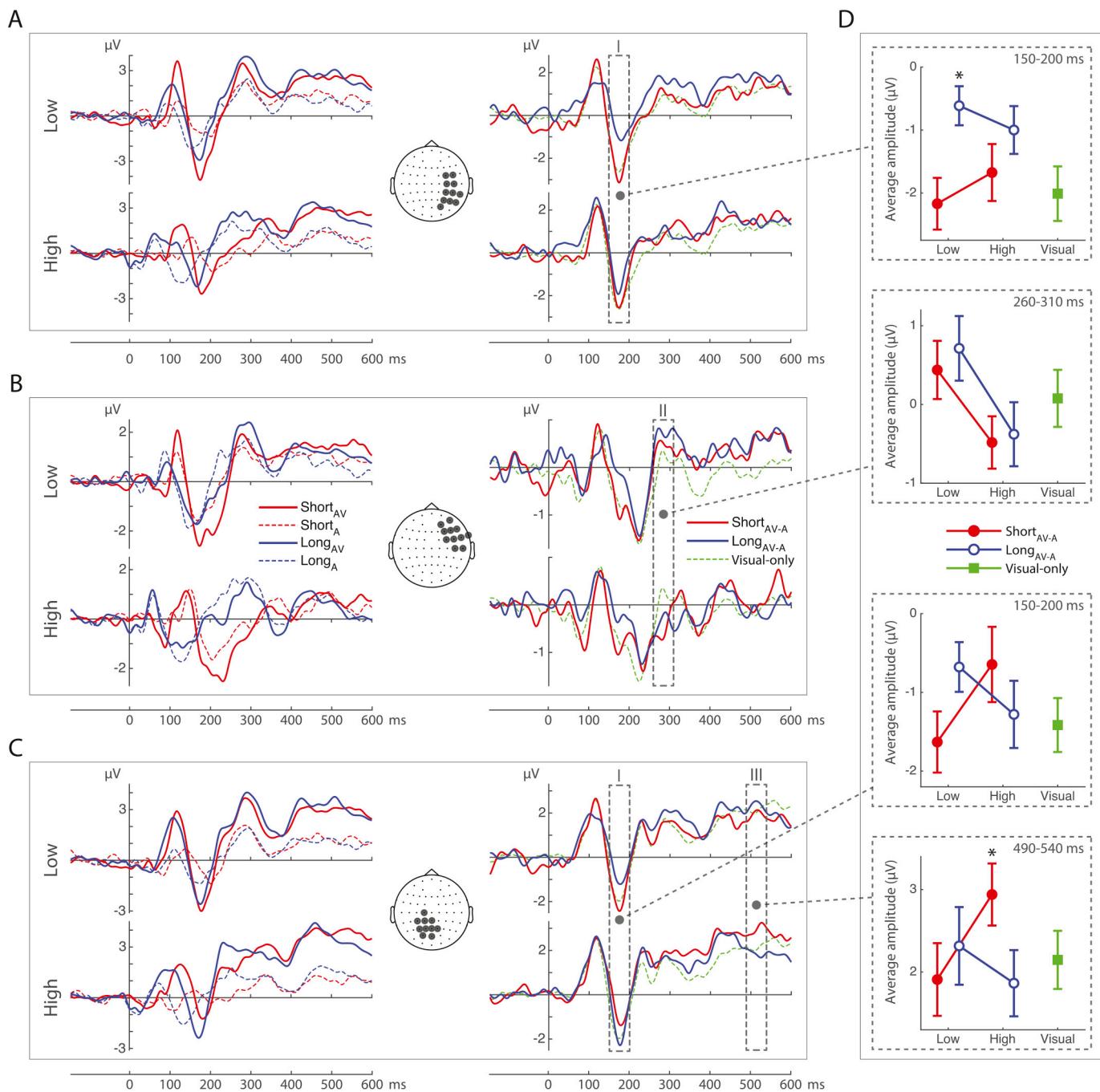
We also computed average potentials within the identified time windows for three exemplar sites and these values are indicated by Fig. 4. Examination of scalp topographies revealed that time interval effects were dominant over right centro-parietal electrodes within the 150–200 ms range (Fig. 4A). This is confirmed by the repeated-measures ANOVA test (time interval and disparity level as factors) on the average potentials of difference ERPs (Fig. 4D). In this time-range, we found a significant effect of the time interval, but the effect of disparity and two-way interaction were not significant (Table 1). The average potentials for short time intervals were smaller than those of long intervals. Moreover, we compared the average potential of a specific timing condition (AV-A) with that of the visual-only (V) condition. For the long time interval condition of low disparity, the average potential was significantly higher than that of visual-only. This suggests a super-additive ( $AV > A + V$ ) audiovisual interaction for this condition.

The disparity level became influential over frontal sites and was most dominant at the right hemisphere (Fig. 4B). Additional ANOVA tests on the average activities of difference ERPs indicated a significant effect of disparity for 260–310 ms time range. For this time range, an increase in disparity level led to a decrease in the average activity. On the other hand, there was no significant effect of time interval and no two-way interaction. In terms of these average amplitudes (AV-A), comparison of each condition with the visual-only condition did not indicate a significant difference, suggesting that there was no super- or sub-additive interaction within this time frame over these electrodes (Fig. 4D).

Our behavioral results indicated a significant increase in the perceived speed difference between short and long conditions as the disparity level was increased. In a previous study, Kafaligonul and Stoner (2010) found similar results. They also explicitly showed that when the temporal disparity is increased, the perceived speed of the short and long conditions increase and decrease, respectively. Overall, these behavioral results suggest a significant interaction between time interval and disparity level. Therefore, the cluster of electrodes, over which a significant two-way interaction is observed, has particular importance here. The averaged ERPs from an exemplar site for the two-way interaction are shown in Fig. 4C. At this exemplar site, the interaction between time



**Fig. 3. Scalp topographies and source estimations.** The significant spatiotemporal clusters were present in three distinct time windows (I: 150–200 ms; II: 260–310 ms; III: 490–540 ms). The temporal locations of these time windows are marked on the timeline. (A) Difference topographical maps and whole-brain t-maps from sLORETA source estimations within each identified time window are displayed in different panels separated by dashed borders. For the main effects and two-way interaction, derived waveforms were used (Time Interval: combined short<sub>AV-A</sub> – combined long<sub>AV-A</sub>; Disparity Level: combined low<sub>AV-A</sub> – combined high<sub>AV-A</sub>; Interaction: subtracted low<sub>AV-A</sub> – subtracted high<sub>AV-A</sub>; see Methods for additional information). The electrodes, which were part of a significant spatiotemporal cluster for at least 20 ms of contiguous data in the time-range, are marked by filled circles on the difference maps. Viewing angle of the 3D inflated brain templates was arranged according to the topographical maps and the significant scalp sites above. Color bar under each 3D brain template represents voxel t-values. The sign of the difference between derived waveforms is represented by negative (blue) and positive (red) t-values. Scaling was arranged so that shaded colors indicate extreme t-values (i.e., upper/lower limits marked on the color bar and beyond). (B) Voltage topographical maps of individual auditory timing conditions. In each panel, the topographical map of difference (AV-A) ERP for each time interval and disparity level condition is shown in separate columns and rows, respectively.



**Fig. 4. Averaged activities from three exemplar scalp sites ( $n = 19$ ).** The exemplar sites for the time interval and disparity level consisted of all the electrodes highlighted in Fig. 3. The exemplar site for the two-way interaction included electrodes which were part of both early and late significant spatiotemporal cluster. (A) Right centro-parietal scalp sites (PO4, P4, P6, P8, CP4, CP6, TP8, C4, C6, FC4, FC6), (B) Right frontal scalp sites (C6, T8, FC4, FC6, FT8, FT10, F4, F6, F8, AF4, AF8), (C) Medial parietal scalp sites (PO3, POz, P3, P1, Pz, P2, CP3, CP1, CPz, C1). Grand-averaged ERPs for audiovisual and auditory-only stimulation are shown in the left plots. The corresponding difference (AV-A) and visual-only ERPs are displayed in the right plots. The ERPs for low and high disparity level conditions are shown in separate rows. (D) Averaged ERP amplitudes in different time windows as a function of temporal disparity level. The temporal location of each time window is shown in the difference ERP plots on the left. The values for each time window and each exemplar site are displayed in separate plots. In each plot, the filled (red) and open (blue) circles correspond to short and long time interval conditions, respectively. The green square indicates the mean value for the visual-only condition. Error bars indicate standard error ( $\pm$ SEM) across participants. A significant difference between each auditory timing and the visual-only condition was marked with an asterisk sign (two-tailed paired  $t$ -test,  $p < 0.05$ ).

interval and disparity level was mainly in the 150–200 ms and 490–540 ms ranges. In the 150–200 ms time range, the time interval had distinct effects on the average potentials of difference ERPs for each disparity level (Fig. 4D). For the low disparity level, the average potential of the short condition was smaller than that of the long condition. It was in the reverse direction for the high disparity level. As opposed to

changes in the percentage values of behavioral data, the difference between average values of the time interval conditions was larger in the low disparity level. Post-hoc pairwise comparisons indicated that the difference between short and long time intervals for the low level was significant ( $t_{18} = -3.385, p = 0.003$ ), but the difference was not significant for the high level ( $t_{18} = 1.489, p = 0.154$ ). For the 490–540 ms time range,

**Table 1**

**Two-way repeated measures ANOVAs on the averaged ERP amplitudes.** The table summarizes the results of ANOVA tests on the data shown in Fig. 4D. In each row, the values for each time window are shown. The values of each exemplar site are also grouped in separate rows. Significant *p* values (*p* < 0.05) are highlighted in bold.

	Time Interval			Disparity Level			Time Interval x Disparity Level		
	F <sub>1,18</sub>	<i>p</i>	$\eta_p^2$	F <sub>1,18</sub>	<i>p</i>	$\eta_p^2$	F <sub>1,18</sub>	<i>p</i>	$\eta_p^2$
<b>Exemplar Site 1: Fig. 4A</b>									
150–200 ms	10.361	<b>0.006</b>	0.365	0.039	0.846	0.002	1.515	0.234	0.078
<b>Exemplar Site 2: Fig. 4B</b>									
260–310 ms	0.385	0.543	0.021	18.190	< 0.001	0.503	0.138	0.715	0.008
<b>Exemplar Site 3: Fig. 4C</b>									
150–200 ms	0.541	0.417	0.029	0.317	0.580	0.017	7.557	<b>0.013</b>	0.296
490–540 ms	0.866	0.364	0.046	0.832	0.364	0.046	10.641	<b>0.004</b>	0.372

the time interval effects on the difference (AV-A) ERPs were also distinct and in the opposite direction. In line with the change in percentage values, the difference between average potentials for short and long time interval conditions was larger in the high disparity level. Pairwise comparisons indicated that the effect of time interval was only significant for the high disparity level ( $t_{18} = 2.463$ ,  $p = 0.024$ ), such that the long time interval yielded higher potentials than the short interval. In terms of changes relative to the visual-only condition, we only found a super-additive interaction for the short time interval of the high disparity level in the 490–540 ms time window. For all the exemplar sites, we also carried out statistical analyses in the –145 ms to 0 ms time range. These analyses did not reveal any significant main effect or two-way interaction.

#### 4. Discussion

Although it has been well known that auditory timing through concurrent stimulation affects visual motion perception, there is a limited understanding of when and where in the brain the underlying processes operate. Our study provides the first systematic EEG investigation on this matter. Using a speed judgment paradigm, we identified three distinct clusters of electrodes over which auditory timing takes place. Over medial parietal and parieto-occipital sites, we found a significant interaction between auditory time interval (short vs. long) and disparity level (low vs. high) in the 150–200 ms and 490–540 ms time windows. In the 150–200 ms time range (N1 component), the main effect of time interval was also found to be significant. But, this time interval effect was mainly located over right centro-parietal, parietal, and temporal scalp sites. The main effect of disparity level was dominant over (right) frontal regions within the 260–310 ms time range. In particular, the disparity level was designed to manipulate the predictive power of clicks. Compared to time interval, it was expected to engage auditory modulations at later stages of motion processing. Therefore, our results fit well with this original hypothesis and prediction. Our findings also indicate the involvement of various processes that take place at different stages of visual motion processing. Thus, they also support the general hypothesis that audio-visual integration involves cortical networks operating at distinct stages of sensory processing. In what follows, we discuss each of these identified scalp sites within the context of audiovisual temporal processing and recent findings on the multisensory nature of motion perception. At the final part, we also evaluate our findings from a general perspective and provide their implications for multisensory research.

##### 4.1. Early and late auditory timing effects over medial parietal scalp sites

We found significant modulations by auditory timing (i.e., time interval and disparity level interactions) centered over the medial parietal electrodes in early (150–200 ms) and late (490–540) ERP components. These changes were also present at some of the centro-parietal, parieto-occipital, and occipital electrodes. Previous multisensory studies have consistently reported significant audiovisual interactions within the N1 component range (e.g., Molholm et al., 2002; Teder-Sälejärvi et al., 2002,

2005). These interactions have been mostly interpreted as the influence of auditory inputs on the sensory processing in visual cortex. Of note, Stekelenburg and Vroomen (2005) examined the influence of a single auditory click on the timing of a flashed visual object using a flash-lag paradigm. As in the original temporal ventriloquism paradigm, they found that auditory timing significantly affected perceived visual timing. The amplitudes of the N1 component over parieto-occipital sites were found to be significantly modulated and these modulations were suggested to correlate with the magnitude of the changes in perceived visual timing. More recently, Zhao et al. (2018) have observed significant super-additive interactions in this time-range that contribute to stream/bounce illusion in which a transient static sound changes the perceived motion of two moving objects. Overall, our findings are in agreement with these results by showing significant auditory modulations in the N1 component range (see also below, for the time interval effect). On the other hand, the later modulations were mostly in agreement with perceptual performance. Within this time window (490–540 ms), as in perceptual performance, the difference between short and long time interval ERPs (i.e., AV-A ERPs) increased when the disparity level became higher. The short time interval of high disparity level also led to a significant super-additive interaction in this range. Therefore, in an apparent motion design based on pairs of clicks and motion frames, our findings emphasize the role of the late ERP components over parietal and parieto-occipital regions.

Similar to our observations here, it was previously pointed out that activities over parieto-occipital and occipital regions were enhanced while subjects were discriminating visual motion directions and listening to acoustic noise in tandem (Gleiss and Kayser, 2014; Kayser et al., 2017). In these EEG studies, random-dot displays were used as visual motion and the acoustic noise was either static or also carried out direction information. In particular, when visual and auditory motion direction were congruent in terms of motion direction, the perceptual performance was high (Kayser et al., 2017). It should also be noted that static noise may also increase performance (Kim et al., 2012; Gleiss and Kayser, 2014). Further investigation of the neural activity revealed multiple processes operating at different time scales. However, similar to our observations here, the auditory modulations in later components over occipital regions were found to be particularly relevant to the enhancement in perceptual performance. These results have been interpreted to mean that the later modulations over these regions may be dependent on feedback from higher association areas which guide multisensory influences based on task requirements.

##### 4.2. Auditory time intervals and right centro-parietal scalp sites

In the present study, the differences between the averaged ERP amplitudes (AV-A) of short and long auditory time intervals were mainly significant over right centro-parietal and parietal sites in the N1 component (150–200 ms) range. As in similar multisensory studies (e.g., Zhao et al., 2018), a significant super-additive interaction in the long time interval of low disparity level was also observed. Recently, Kaya et al. (2017) examined the effects of auditory time interval adaptation on

the ERPs elicited by apparent motion. They found significant aftereffects (i.e., short vs. long time interval adaptation) on the late ( $>300$  ms) components over right parietal and centro-parietal sites. Though the time windows do not exactly match, these findings from concurrent stimulation and adaptation design suggest that auditory time intervals play an important role in the neural activity and in the motion processing over right parietal and centro-parietal cortices. In fact, this is consistent with previous studies emphasizing the role of parietal areas in motion estimation. Cortical areas over this region have been known to be multi-sensory and to play important roles in motion processing. For instance, right inferior parietal lobe (rIPL) becomes significantly active when subjects perceive an apparent motion. Moreover, right IPL can be selectively activated by the visual motion engaging high-level attention-based motion processing, suggesting that this region represents a further stage than low-level cortical areas [e.g., areas V1 (primary visual cortex), V3A and MT] in the visual motion hierarchy (Claeys et al., 2003). Mounting evidence also suggests the involvement of right parietal cortices in the processing of sub-second time intervals (Battelli et al., 2007; Shuler, 2016). Right parietal lesions (e.g., right IPL) specifically degraded and introduced deficits in timing tasks (Battelli et al., 2001, 2003). The activity in the right parietal cortex (e.g., rIPL) can be modulated through duration adaptation (Hayashi et al., 2015). Moreover, an application of brain stimulation to right parietal cortices resulted in changes specific to sub-second duration judgments (Bueti et al., 2008; Dormal et al., 2016). Of particular relevance to the current study, Bueti et al. (2008) showed that the disruption of the right parietal cortex interfered with both auditory and visual time perception. This suggests that right parietal cortex may have an important role in perceptual tasks highly dependent on the timing of stimulation. Accordingly, in combination with these studies, our findings within the context of motion processing highlight the importance of the right parietal regions in audiovisual temporal processing. It is important to note that supplementary source estimations indicated additional neural generators in insula and the auditory specific regions of temporal lobe. Using different audiovisual motion paradigms, previous research showed audiovisual interactions in these regions (Lewis et al., 2000; Senkowski et al., 2007; Getzmann and Lewald, 2014). Interestingly, a task requiring cross-modal speed comparison of auditory and visual motions revealed significant enhancement of anterior insula activity (Lewis et al., 2000).

#### 4.3. Temporal disparity level and (right) frontal scalp sites

In addition to audiovisual interactions at low-level cortical areas, previous EEG investigations revealed interactions over fronto-central and frontal electrodes with basic audiovisual inputs including simultaneously presented brief sound and visual flash (Molholm et al., 2002). Similar audiovisual interactions over these scalp sites have been also indicated by studies using more naturalistic sounds and images (Senkowski et al., 2007; Stekelenburg and Vroomen, 2007). The late audiovisual interactions over these regions were mostly found to be context dependent such that its amount significantly changed based on the mismatch between the content of naturalistic sound and image. More recently, the ERP findings by Cecere et al. (2017) have suggested that the interactions over these regions may also depend on the temporal properties of audiovisual stimulation. Through a relatively complicated design in terms of temporal dynamics of stimulation, our results here extend previous findings by showing that the temporal disparity level between the auditory time intervals (and also the disparity level between each click and each motion frame/visual flash) can become significant over fronto-central and frontal electrodes. While our results support the modulation of frontal activities by the temporal properties of audiovisual stimulation in general, the time-range and frontal electrode locations (i.e., the spatiotemporal profile of the modulations) do not exactly match with those reported by previous studies. The significant effects of disparity level over these regions are also meaningful in terms of motion processing. In particular, the prefrontal cortex (PFC) is generally

associated with high-level sensory processing and neurons selective to the direction and speed of visual motion have been observed in this region (Zaksas and Pasternak, 2006; Hussar and Pasternak, 2013; Wimmer et al., 2016). Of particular relevance to the experimental design here, it has been shown that many PFC neurons can be selectively engaged in discrimination tasks which require subjects to compare the directions or speeds of two sequentially presented visual motion. The activations over this region have been proposed to subserve the comparison of sensory signals and to play an important role in perceptual tasks during which the comparison of the remembered and current stimuli has to be performed. Furthermore, subdivisions of this area (e.g., Brodmann area 8) are known to receive inputs from both motion pathway and auditory cortex (Romanski, 2007) and the PFC has been proposed to play a role in audiovisual motion integration recently (Chaplin et al., 2018).

As mentioned in the previous sections, the interaction between disparity level and time interval became significant in a later ERP component of medial parietal and parieto-occipital regions. Moreover, this late (490–540 ms) two-way interaction was mostly in agreement with perceptual performance in terms of changes due to disparity level increments. The time-course of these auditory modulations may be achieved through top-down projections originating from frontal regions. Such cortical mechanisms may indirectly (or directly) influence the activities over relatively low-level sensory areas and thus be involved in shaping perception (e.g., Knight et al., 1999; Miller and D'Esposito, 2005; Zhang et al., 2014). Specifically, we suggest that the activations over frontal regions, due to a change in the temporal disparity level (and hence due to the predictive power of clicks), may subsequently modulate the neural activities in parietal and parieto-occipital regions. Likewise, the right centro-parietal regions may influence these regions through feedback connections.

Another interesting point worthwhile to mention is that the significant modulations over frontal areas might be due to just decisional/response biases on auditory time intervals. In other words, as the disparity level was increased, the difference between two auditory time intervals became more obvious and observers could have only used this auditory feature rather than performing speed discrimination task on visual apparent motion. Thus, this change could have led to significant differences between the ERPs of low and high disparity level conditions over frontal regions. While we consider that this is unlikely due to the procedure applied and the behavioral results of catch trials, our ERP results are also informative to evaluate this interpretation. Such interpretation also implies that observers may have used a different judgment strategy for performing the task in bimodal (AV) conditions and these conditions may have also required different task demands than the visual-only (V) condition. Therefore, according to such interpretation, all of the bimodal difference ERPs should have been significantly above or below that of visual-only. However, as shown by Fig. 4, this is not the case at the frontal exemplar site. In particular, the difference ERPs of both high disparity level conditions were not significantly different than visual-only ERP (Fig. 4D). The ERPs from other exemplar sites point out a similar situation. Overall, our ERP results suggest that the contribution of any response/decisional bias to the identified time windows and exemplar sites was limited.

#### 4.4. Motion estimation as a multisensory and a multistage process

Motion and speed estimation have been extensively investigated by vision scientists (Kolers, 1972; Nakayama, 1985; Burr and Thompson, 2011). Much of the previous work on speed estimation has been performed in area MT+ [the presumed human homolog of area MT and MST (medial superior temporal area)], since this area is preferentially activated during speed discrimination tasks (Huk and Heeger, 2000) and speed discrimination of non-human primates in which area MT has been lesioned is impaired (Newsome and Paré, 1988; Orban et al., 1995; Rudolph and Pasternak, 1999). Also, the microstimulation of area MT alters speed perception (Liu and Newsome, 2005). In these studies, the

manipulations were only based on visual parameters and they were mostly focused on speed computations in low-level motion areas (e.g., area MT). That is to say, they were restricted to the modality of vision and specific cortical areas. On the other hand, multisensory research ushered a new perspective of motion perception, wherein information provided by other modalities (e.g., audition) is also involved in motion and speed computations (Soto-Faraco et al., 2003; Hidaka et al., 2015). In agreement with this perspective, our findings here highlight the multisensory nature of motion processing. Using an important aspect of motion (i.e., speed), we showed that evoked activity to apparent motion and perception significantly change by the timing information provided through audition. Specifically, our EEG study can be viewed as a novel extension of previous studies on speed. By pointing out specific scalp sites and ERP components, our findings (within the context of an audiovisual paradigm) provide comprehensive information on how information provided by other modalities are involved in speed estimation.

As mentioned in the previous sections, the experimental factors (time interval and disparity level) defined based on auditory timing had influences over distinct cluster of electrodes and ERP components. Moreover, the interaction between these factors time interval and disparity level became significant over another cluster of electrodes in early and late neural activities. By indicating that auditory timing takes place at various stages of motion processing, our findings here fit well with the notion that multisensory integration is a multistage process (Cecere et al., 2017). We consider that the effects of auditory timing on perceived speed may be achieved through the dynamic interplay between these identified regions at different stages of cortical processing. From this perspective, the present findings also reveal the dynamic and highly interactive nature of multisensory processing. Interestingly, previous reports on different audiovisual motion paradigms have also emphasized the involvement of various cortical areas at different stages of sensory processing (Baumann and Greenlee, 2007; Lewis and Noppeney, 2010). They further suggested that each cortical area may have a specific functional role in motion perception and multisensory processing.

#### 4.5. Limitations and future directions

In our EEG study, we used temporally centered apparent motion and click sequences. When the temporal disparity level was changed between the two auditory time interval conditions, the relative timing between each motion frame and click was also changed. That is to say, we restricted ourselves to certain temporal profiles in this basic design. Moreover, an emerging view suggests that multisensory integration is flexible, context-dependent and that behavioral goals determine the use of integration mechanisms dynamically (van Atteveldt et al., 2014). A challenge for the future is to characterize the identified scalp sites under a rich repertoire of audiovisual temporal profiles and different perceptual tasks (e.g., time interval estimation). Our experimental design was oriented to test the violation of the additive model and to reveal the modulations of nonlinear components. To achieve this aim, we had to include both bimodal and unimodal conditions for each temporal factor. An increase in the number of conditions led to a decrease in the number of trials per condition. Also, this approach (i.e., testing the additive model) required us to perform analyses on the derived waveforms. Using the derived waveforms with these number of trials may have some limitations to carry out standard source localization analyses. Therefore, we provided the outcome of these analyses as additional supporting information. Further detailed investigation of the auditory timing effects in the anatomical space will be informative.

#### 5. Conclusion

In conclusion, the present study highlights the multisensory nature of motion estimation and also provides specific information on the cortical processes underlying this aspect of motion processing. Using a speed judgment paradigm, we found distinct cortical regions over which

auditory timing takes place for visual motion. More specifically, our analyses on the spatiotemporal profile of the neural activity point to the involvement of different mechanisms operating at distinct stages of motion processing. Accordingly, these findings, in conjunction with a variety of related converging evidence, demonstrate the diversified and dynamic nature of audiovisual temporal processing.

#### Conflicts of interest

The authors declare no competing financial interests.

#### Acknowledgments

We thank Can Oluk for technical assistance in data collection. We are also grateful to Aaron Clarke, Jenny Ball and Buse M. Urgen for discussions and comments on the manuscript. This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK Grant 113K547).

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2019.05.062>.

#### References

- Allen, P.J., Polizzi, G., Krakow, K., Fish, D.R., Lemieux, L., 1998. Identification of EEG events in the MR scanner: the problem of pulse artifact and a method for its subtraction. *Neuroimage* 8, 229–239.
- Battelli, L., Cavanagh, P., Intriligator, J., Tramo, M.J., Hénaff, M.A., Michél, F., Barton, J.J., 2001. Unilateral right parietal damage leads to bilateral deficit for high-level motion. *Neuron* 32, 985–995.
- Battelli, L., Cavanagh, P., Martini, P., Barton, J.J., 2003. Bilateral deficits of transient visual attention in right parietal patients. *Brain* 126, 2164–2174.
- Battelli, L., Pascual-Leone, A., Cavanagh, P., 2007. The ‘when’ pathway of the right parietal lobe. *Trends Cognit. Sci.* 11, 204–210.
- Baumann, O., Greenlee, M.W., 2007. Neural correlates of coherent audiovisual motion perception. *Cerebr. Cortex* 17, 1433–1443.
- Besle, J., Fort, A., Giard, M., 2004. Interest and validity of the additive model in electrophysiological studies of multisensory interactions. *Cogn. Process.* 5, 189–192.
- Brainard, D., 1997. The psychophysics toolbox. *Spatial Vis.* 10, 433–436.
- Bueti, D., Bahrami, B., Walsh, V., 2008. Sensory and association cortex in time perception. *J. Cogn. Neurosci.* 20, 1054–1062.
- Burr, D., Banks, M., Morrone, M., 2009. Auditory dominance over vision in the perception of interval duration. *Exp. Brain Res.* 198, 49–57.
- Burr, D., Thompson, P., 2011. Motion psychophysics: 1985–2010. *Vis. Res.* 51, 1431–1456.
- Cappe, C., Thut, G., Romei, V., Murray, M.M., 2010. Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *J. Neurosci.* 30, 12572–12580.
- Cecere, R., Gross, J., Thut, G., 2016. Behavioural evidence for separate mechanisms of audiovisual temporal binding as a function of leading sensory modality. *Eur. J. Neurosci.* 43, 1561–1568.
- Cecere, R., Gross, J., Willis, A., Thut, G., 2017. Being first matters: topographical representational similarity analysis of ERP signals reveals separate networks for audiovisual temporal binding depending on the leading sense. *J. Neurosci.* 37, 5274–5287.
- Chaplin, T.A., Rosa, M.G.P., Lui, L.L., 2018. Auditory and visual motion processing and integration in the primate cerebral cortex. *Front. Neural Circuits* 12, 93.
- Chen, L., Vroomen, J., 2013. Intersensory binding across space and time: a tutorial review. *Atten. Percept. Psychophys.* 75, 790–811.
- Claeys, K.G., Lindsey, D.T., De Schutter, E., Orban, G.A., 2003. A higher order motion region in human inferior parietal lobule: evidence from fMRI. *Neuron* 40, 451–452.
- Dormal, V., Javadi, A.H., Pesenti, M., Walsh, V., Cappelletti, M., 2016. Enhancing duration processing with parietal brain stimulation. *Neuropsychologia* 85, 272–277.
- Fendrich, R., Corballis, P.M., 2001. The temporal cross-capture of audition and vision. *Percept. Psychophys.* 63, 719–725.
- 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.
- Getzmann, S., Lewald, J., 2014. Modulation of auditory motion processing by visual motion: early crossmodal interactions in human auditory cortices. *J. Psychophysiol.* 28, 82–100.
- Gleiss, S., Kayser, C., 2014. Oscillatory mechanisms underlying the enhancement of visual motion perception by multisensory congruency. *Neuropsychologia* 53, 84–93.
- Groppe, D.M., Urbach, T.P., Kutas, M., 2011. Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. *Psychophysiology* 48, 1711–1725.

- Hayashi, M.J., Ditye, T., Harada, T., Hashiguchi, M., Sadato, N., Carlson, S., Walsh, V., Kanai, R., 2015. Time adaptation shows duration selectivity in the human parietal cortex. *PLoS Biol.* 13, e1002262.
- 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.
- Hoffmann, S., Labrenz, F., Themann, M., Wascher, E., Beste, C., 2014. Crosslinking EEG time-frequency decomposition and fMRI in error monitoring. *Brain Struct. Funct.* 219, 595–605.
- Huk, A.C., Heeger, D.J., 2000. Task-related modulation of visual cortex. *J. Neurophysiol.* 83, 3525–3536.
- Hussar, C.R., Pasternak, T., 2013. Common rules guide comparison of speed and direction of motion in the dorsolateral prefrontal cortex. *J. Neurosci.* 33, 972–986.
- Kafaligonul, H., Albright, T.D., Stoner, G.R., 2018. Auditory modulation of spiking activity and local field potentials in area MT does not appear to underlie an audiovisual temporal illusion. *J. Neurophysiol.* 120, 1340–1355.
- Kafaligonul, H., Stoner, G.R., 2010. Auditory modulation of visual apparent motion with short spatial and temporal intervals. *J. Vis.* 10 (12), 31.
- Kafaligonul, H., Stoner, G.R., 2012. Static sound timing alters sensitivity to low-level visual motion. *J. Vis.* 12 (11), 2.
- 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. *Eur. J. Neurosci.* 46, 2325–2338.
- Kayser, S.J., Philiastides, M.G., Kayser, C., 2017. Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations. *Neuroimage* 148, 31–41.
- Kim, R., Peters, M.A., Shams, L., 2012. 0 + 1 > 1: how adding noninformative sound improves performance on a visual task. *Psychol. Sci.* 23, 6–12.
- Knight, R.T., Staines, W.R., Swick, D., Chao, L.L., 1999. Prefrontal cortex regulates inhibition and excitation in distributed neural networks. *Acta Psychol.* 101, 159–178.
- Kolers, P.A., 1972. Aspects of Motion Perception. Pergamon, Oxford.
- Lewis, J.W., Beauchamp, M.S., DeYoe, E.A., 2000. A comparison of visual and auditory motion processing in human cerebral cortex. *Cerebr. Cortex* 10, 873–888.
- Lewis, R., Noppeney, U., 2010. Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *J. Neurosci.* 30, 12329–12339.
- Liu, J., Newsome, W.T., 2005. Correlation between speed perception and neural activity in the middle temporal visual area. *J. Neurosci.* 25, 711–722.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164 (1), 177–190.
- Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Boomsma, D., Cannon, T., Kawashima, R., Mazoyer, B., 2001. A probabilistic atlas and reference system for the human brain: international consortium for brain mapping (ICBM). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 1293–1322.
- Miller, B.T., D'Esposito, M., 2005. Searching for “the top” in top-down control. *Neuron* 48, 535–538.
- Mishra, J., Martinez, A., Sejnowski, T.J., Hillyard, S.A., 2007. Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *J. Neurosci.* 27, 4120–4131.
- Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., Foxe, J.J., 2002. Multisensory auditory-visual interactions during early sensory processing in humans: a high density electrical mapping study. *Brain Res. Cogn. Brain Res.* 14, 115–128.
- Morein-Zamir, S., Soto-Faraco, S., Kingstone, A., 2003. Auditory capture of vision: examining temporal ventriloquism. *Brain Res. Cogn. Brain Res.* 17, 154–163.
- Murray, M.M., Lewkowicz, D.J., Amedi, A., Wallace, M.T., 2016. Multisensory processes: a balancing act across the lifespan. *Trends Neurosci.* 39, 567–579.
- Nakayama, K., 1985. Biological image motion processing: a review. *Vis. Res.* 25, 625–660.
- Naue, N., Rach, S., Strüber, D., Huster, R.J., Zehle, T., Körner, U., Herrmann, C.S., 2011. Auditory event-related response in visual cortex modulates subsequent visual responses in humans. *J. Neurosci.* 31, 7729–7736.
- Newsome, W.T., Paré, E.B., 1988. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci.* 8, 2201–2211.
- Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* 15, 1–25.
- Ogulmus, C., Karacaoglu, M., Kafaligonul, H., 2018. Temporal ventriloquism along the path of apparent motion: speed perception under different spatial grouping principles. *Exp. Brain Res.* 236, 629–643.
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 156869.
- Orban, G.A., Saunders, R.C., Vandebussche, E., 1995. Lesions of the superior temporal cortical motion areas impair speed discrimination in the macaque monkey. *Eur. J. Neurosci.* 7, 2261–2276.
- Orchard-Mills, E., Van Der Burg, E., Alais, D., 2016. Crossmodal correspondence between auditory pitch and visual elevation affects temporal ventriloquism. *Perception* 45, 409–424.
- Pascual-Marqui, R.D., 2002. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find. Exp. Clin. Pharmacol.* 24 (Suppl. D), 5–12.
- Pelli, D., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vis.* 10, 437–442.
- Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neurophysiol.* 72, 184–187.
- Recanzone, G.H., 2003. Auditory influences on visual temporal rate perception. *J. Neurophysiol.* 89, 1078–1093.
- Romanski, L.M., 2007. Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. *Cerebr. Cortex* 17, 61–69.
- Rudolph, K., Pasternak, T., 1999. Transient and permanent deficits in motion perception after lesions of cortical areas MT and MST in the macaque monkey. *Cerebr. Cortex* 9, 90–100.
- Sekihiwa, K., Sahani, M., Nagarajan, S.S., 2005. Localization bias and spatial resolution of adaptive and non-adaptive spatial filters for MEG source reconstruction. *Neuroimage* 25, 1056–1067.
- Senkowski, D., Saint-Amour, D., Kelly, S.P., Foxe, J.J., 2007. Multisensory processing of naturalistic objects in motion: a high-density electrical mapping and source estimation study. *Neuroimage* 36, 877–888.
- Shi, Z., Chen, L., Müller, H.J., 2010. Auditory temporal modulation of the visual Ternus effect: the influence of time interval. *Exp. Brain Res.* 203, 723–735.
- Shuler, M.G., 2016. Timing in the visual cortex and its investigation. *Curr. Opin. Behav. Sci.* 8, 73–77.
- Soto-Faraco, S., Kingstone, A., Spence, C., 2003. Multisensory contributions to the perception of motion. *Neuropsychologia* 41, 1847–1862.
- Spence, C., 2011. Crossmodal correspondences: a tutorial review. *Atten. Percept. Psychophys.* 73, 971–995.
- Stekelenburg, J.J., Vroomen, J., 2005. An event-related potential investigation of the time-course of temporal ventriloquism. *Neuroreport* 16, 641–644.
- Stekelenburg, J.J., Vroomen, J., 2007. Neural correlates of multisensory integration of ecologically valid audiovisual events. *J. Cogn. Neurosci.* 19, 1964–1973.
- Stevenson, R.A., Ghose, D., Fister, J.K., Sarko, D.K., Altieri, N.A., Nidiffer, A.R., Wallace, M.T., 2014. Identifying and quantifying multisensory integration: a tutorial review. *Brain Topogr.* 27, 707–730.
- Talsma, D., Senkowski, D., Soto-Faraco, S., Woldorff, M.G., 2010. The multifaceted interplay between attention and multisensory integration. *Trends Cognit. Sci.* 14, 400–410.
- Teder-Sälejärvi, W.A., Di Russo, F., McDonald, J.J., Hillyard, S.A., 2005. Effects of spatial congruity on audio-visual multimodal integration. *J. Cogn. Neurosci.* 17 (9), 1396–1409.
- Teder-Sälejärvi, W.A., McDonald, J.J., Di Russo, F., Hillyard, S.A., 2002. An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Brain Res. Cogn. Brain Res.* 14, 106–114.
- Thorne, J.D., Debener, S., 2014. Look now and hear what's coming: on the functional role of cross-modal phase reset. *Hear. Res.* 307, 144–152.
- Urgen, B.A., Kutas, M., Saygin, A.P., 2018. Uncanny valley as a window into predictive processing in the social brain. *Neuropsychologia* 114, 181–185.
- van Atteveldt, N., Murray, M.M., Thut, G., Schroeder, C.E., 2014. Multisensory integration: flexible use of general operations. *Neuron* 81, 1240–1253.
- Vroomen, J., Keetels, M., 2006. The spatial constraint in intersensory pairing: no role in temporal ventriloquism. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 1063–1071.
- Welch, R.B., Warren, D.H., 1980. Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.* 88, 638–667.
- Wimmer, K., Spinelli, P., Pasternak, T., 2016. Prefrontal neurons represent motion signals from across the visual field but for memory-guided comparisons depend on neurons providing these signals. *J. Neurosci.* 36, 9351–9364.
- World Medical Association, 2013. Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 310 (20), 2191–2194.
- Zaksas, D., Pasternak, T., 2006. Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *J. Neurosci.* 26, 11726–11742.
- Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J.P., Chang, W.C., Jenvey, S., Miyamichi, K., Luo, L., Dan, Y., 2014. Long-range and local circuits for top-down modulation of visual cortex processing. *Science* 345, 660–665.
- Zhao, S., Wang, Y., Xu, H., Feng, C., Feng, W., 2018. Early cross-modal interactions underlie the audiovisual bounce-inducing effect. *Neuroimage* 174, 208–218.