DR. UTKU KAYA (Orcid ID : 0000-0001-6127-2761)

DR. HULUSI KAFALIGONUL (Orcid ID : 0000-0001-5033-4138)

Article type : Original article

Audiovisual Interactions in Speeded Discrimination of a Visual Event

Utku Kaya[1,2,3], Hulusi Kafaligonul[1,4*]

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

[2]Informatics Institute, Middle East Technical University, Ankara, Turkey

[3]Department of Anesthesiology, University of Michigan, Ann Arbor, MI

[4]Interdisciplinary Neuroscience Program, Aysel Sabuncu Brain Research Center, Bilkent University, Ankara, Turkey

Running Title: Audiovisual interactions and reaction times

Funding: The Scientific and Technological Research Council of Turkey (grant number 113K547) and the Turkish Academy of Sciences (TUBA-GEBIP Award).

*Correspondence:
Hulusi Kafaligonul
Aysel Sabuncu Brain Research Center
Bilkent University, Ankara 06800, Turkey
Email: hulusi@bilkent.edu.tr
Phone: +90 312 290 3016

Abstract

The integration of information from different senses is central to our perception of the external world. Audiovisual interactions have been particularly well studied in this context and various illusions have been developed to demonstrate strong influences of these interactions on the final percept. Using audiovisual paradigms, previous studies have shown that even task-irrelevant information provided by a secondary modality can change the detection and discrimination of a primary target. These modulations have been found to be significantly dependent on the relative timing between auditory and visual stimuli. Although these interactions in time have been commonly reported, we have still limited understanding of the relationship between the modulations of event-related potentials and final behavioral performance. Here, we aimed to shed light on this important issue by using a speeded discrimination paradigm combined with electroencephalogram. During the experimental sessions, the timing between an auditory click and a visual flash was varied over a wide range of stimulus onset asynchronies and observers were engaged in speeded discrimination of flash location. Behavioral reaction times were significantly changed by click timing. Furthermore, the modulations of evoked activities over medial parietal/parieto-occipital electrodes were associated with this effect. These modulations were within the 126-176 ms time range and more importantly, they were also correlated with the

changes in reaction times. These results provide an important functional link between audiovisual interactions at early stages of sensory processing and reaction times. Together with previous research, they further suggest that early crossmodal interactions play a critical role in perceptual performance.

Keywords: audiovisual interactions, reaction time, visual timing, multisensory, EEG

Introduction

To form a coherent percept of the external world, the brain integrates spatial and temporal information provided by different modalities. Understanding the processes involved in combining information from different sensory modalities has become a focus of research in various areas of neuroscience (Murray & Wallace, 2012; Spence, 2018). Most of the previous studies have been particularly based on auditory and visual modalities. Accordingly, many audiovisual paradigms have been developed to demonstrate the role of crossmodal interactions in sensory processing and final percept (Chen & Vroomen, 2013). Using audiovisual stimulation, previous studies have shown that even task-irrelevant information provided by a secondary modality can change the detection and discrimination of a primary target. Such paradigms have been found to be important for understanding the dynamics of audiovisual interactions at early stages of sensory processing (Zhou, Cheung, & Chan, 2020).

In these studies, simple and brief forms of stimulation (e.g., a click and a visual flash) were typically used. The stimulus onset asynchrony (SOA) between auditory and visual stimuli was varied to understand the nature of audiovisual interactions in the temporal domain. Particularly, the effect of SOA on audiovisual interactions was designed to test the predictions of the phase-

resetting hypothesis. This hypothesis states that events in one sensory modality can reset the phase of oscillations within brain areas specialized for processing another modality (see Thorne & Debener, 2014, for a review). Based on the primary modality (vision or audition), either negative (i.e., SOA ≤ 0) or positive (i.e., SOA ≥ 0) SOAs were used and the sampling rate of SOA values was typically high to test the predictions of phase resetting reliably (e.g., Naue et al., 2011; Thorne, De Vos, Viola, & Debener, 2011). The reaction time (RT) values were found to be significantly dependent on the SOA values and there was a monotonic increase as the absolute value of SOA was increased. More importantly, in the low-frequency oscillations [e.g., electroencephalogram (EEG)], the SOA changed the phase coherency across trials such that only specific SOAs increased coherency as predicted by phase-resetting. A behavioral study (Diederich, Schomburg, & Colonius, 2012) also provides evidence that these changes in phase coherency can be manifested as oscillations (i.e., ripples) on the monotonic increasing trend of RT values from individual subjects.

These findings provide novel and important insights into the nature of audiovisual interactions in time. Through phase-resetting, they first demonstrate how audiovisual interactions can take place over cortical areas that were previously thought to be sensory-specific. Moreover, they reveal that the modulations of low-frequency oscillations over these areas can explain the changes and variations in the final RT values (e.g., Thorne et al., 2011). On the other hand, we have still limited information on the correlation between RT values and changes in the neural activity in terms of event-related potentials (ERPs). Using relatively complex stimulation and perceptual tasks, recent studies suggest the involvement of audiovisual interactions at different stages of sensory processing. For instance, it has been shown that a change in click timing relative to the brief apparent motion frames can take place at both early and late ERP components located over distinct scalp sites (Kaya & Kafaligonul, 2019; Kaya, Yildirim, & Kafaligonul, 2017). Moreover, Cecere, Gross, Willis, and Thut (2017) have found that the temporal order between auditory and visual stimuli is an important factor for engaging audiovisual interactions at distinct scalp sites. In their study, they used a click and a visual flash and systematically varied the timing and the temporal order between these stimuli. Based on the leading modality in time (auditory-leading vs. visual-leading stimulus pairs), they found distinct spatiotemporal maps of EEG activity in terms of audiovisual interactions, suggesting the recruitment of different networks and processes for

evaluating audiovisual synchrony. Their results further support the notion that audiovisual temporal integration may require flexible use of different neural mechanisms (Murray, Lewkowicz, Amedi, & Wallace, 2016; Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010; van Atteveldt, Murray, Thut, & Schroeder, 2014). However, the implications of these findings are not explicitly evaluated within the context of a simple detection or discrimination paradigm. An important question to ask is whether the correlation between RT values and the changes in the spatiotemporal profile of the neural activity is restricted to early ERP components or not.

In the present study, we aimed at understanding the nature of these correlations comprehensively. In particular, we wanted to identify audiovisual interactions at different stages of sensory processing that parallel discrimination performance in terms of RT values. As in previous studies, we used a static click and a visual flash for stimulation and systematically varied the SOA between these stimuli. Critically, our experimental design included both negative (i.e., auditory-leading) and positive (i.e., visual-leading) SOA conditions. Observers were engaged in a speeded discrimination of visual flash location. Building on the recent ERP findings mentioned above, we anticipated on finding audiovisual interactions in both early and late ERP components. Using a relatively complicated audiovisual stimulation and criterion content (e.g., Kaya & Kafaligonul, 2019), previous research suggested the audiovisual interactions in late components are in line with the changes in perceptual performance. Given the recent notion emphasizing that different multisensory processes can be adaptively recruited based on the nature of sensory stimulation and specific task demands (van Atteveldt et al., 2014), the implications of these findings for a simple detection or discrimination task in a wide range of SOAs still remain unclear. Here, using a simple discrimination paradigm, we specifically tested the hypothesis of whether the correlations between RT values and the modulations of neural activity were restricted to late ERP components. Alternatively, as proposed by previous phase-resetting studies, the audiovisual interactions at low-level sensory areas and modulations in early ERP components may play a critical role in shaping final perceptual performance in a simple detection or discrimination paradigm.

Method

Participants

Twenty healthy volunteers (7 females, 19 right-handed, age range of 19-34 years) participated in the study. All participants had normal or corrected-to-normal visual acuity and normal hearing by self-report. None of them reported having a history of neurological disorders. They also gave informed consent before participation. The sample size was commensurate with previous studies using similar settings, audiovisual stimulation, and/or procedure (Kaya & Kafaligonul, 2019; Naue et al., 2011). All procedures were in accordance with the Declaration of Helsinki (World Medical Association, 2013) and approved by the local ethics committee at the School of Medicine, Ankara University.

Apparatus

Stimulus presentation, experimental paradigm, and data acquisition were controlled by MATLAB version 7.12 (The MathWorks, Natick, MA) with the Psychtoolbox 3.0 (Brainard, 1997; Pelli, 1997). Visual stimuli were displayed on a 21-inch CRT monitor (1280 × 1024 pixel resolution, 100 Hz refresh rate) at a viewing distance of 57 cm. A photometer (SpectroCAL, Cambridge Research Systems, Rochester, Kent, UK) was used for luminance calibration and gamma correction of the display. 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 physical 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 the visual stimulus onset. All the experimental sessions were performed in a silent and dimly lit room.

Stimuli and Procedure

As a fixation point, a small red circle (0.3 deg diameter) was presented at the center of the display throughout an experimental block. Visual stimulus was a 50 ms "flashed" bar (0.4 × 3.0 deg with a luminance of 97 cd/m^2) on a gray background (20 cd/m^2). The "flashed" bar was centered 2.5 deg above the central fixation point and presented either 1 deg left or right of the fixation (Figure 1a). A 20 ms "click" (i.e., a brief stationary sound) was used as an auditory stimulus. The click comprised of a rectangular windowed 480 Hz sine-wave carrier and sampled at 44.1 kHz with 8-bit quantization. It was binaurally introduced at 75 dB sound pressure level (Figure 1b). The durations of click and visual flash were exactly the same as those used in our previous study on

apparent motion (Kaya & Kafaligonul, 2019) to have a systematic comparison across findings. The relative timing (SOA) between the visual flash and click were chosen pseudo-randomly from eight values: -160, -120, -80, -40, 0, 40, 80, 120 ms. The negative and positive SOA values corresponded to auditory- and visual-leading conditions, respectively (Figure 1c). The range of SOA values was determined based on pilot behavioral sessions on a few observers. In addition to these bimodal (AV) conditions, two unimodal conditions (auditory-only: A, visual-only: V) were also included in the experiment. Except for presenting either auditory or visual stimulus, the same stimulus parameters of the 0 ms SOA condition were used in these unimodal conditions (see also timelines in Figure 1a-b).

[FIGURE 1 ABOUT HERE]

For each trial, an audiovisual configuration was pseudo-randomly selected from 10 different conditions (8 bimodal and 2 unimodal conditions) and presented according to the timelines in Figure 1. The 600 ms before the visual bar onset was used as a pre-target period. Participants were requested to report the location of the visual bar (left or right, two-alternative forced-choice) via keyboard press as fast as possible (i.e., speeded reaction-time task). Participants were told that the visual bar would be accompanied by a click but to base their responses solely on the visual bar. They were also asked to fixate, passively listen to the click, and not to respond when there was no visual bar during a trial (i.e., auditory-only condition). As soon as the keyboard press, the response was recorded. A trial was ended 850 ms after the onset of the visual bar. The next trial started after a variable inter-trial interval (350-1050 ms). For the auditory-only (A) condition, the timeline of stimulation was exactly the same as that of 0 ms SOA condition with the exception of not displaying the visual bar. As also in bimodal (AV) conditions, observers did not perform any task based on the auditory click in this condition. Our ERP analyses were based on testing the additive model (see *ERP Analyses*, for details). Therefore, when comparing the difference ERPs (AV-A) with that of visual-only (V), major confounding factors (e.g., having no motor response in the difference ERPs) were circumvented through these instructions.

In each experimental block, there were 100 trials (10 conditions x 10 trials per condition). Each participant completed 5 experimental blocks corresponding to a total number of 500 trials (50

trials for each condition). Participants were encouraged to have a short break (approximately less than one minute) between the blocks to maintain high concentration and to prevent fatigue. Prior to these experimental blocks, each participant was also shown examples of the visual and auditory stimuli.

Behavioral Data Analysis

Simple reaction time (RT) has been extensively used to detect changes in the speed of sensory and perceptual processing. As in previous multisensory studies (e.g., Diederich et al., 2012; Navarra, Hartcher-O'Brien, Piazza, & Spence, 2009), we mainly relied on RT values as behavioral measures and thus assessed the perceived timing of a visual event (i.e., flashed bar). The trials in which the location of the visual bar was correctly judged within 150-700 ms range were included in further behavioral and EEG analyses. Based on this criterion, on average only 5.02% of trials per condition (SEM = 0.94%) were excluded. After excluding these and other trials (see *EEG Recording and Preprocessing* for other excluded trials), we calculated average RT values across subjects for each bimodal SOA and visual-only conditions. To determine whether the effect of relative timing between auditory click and visual flash was significant, we applied one-way repeated-measures ANOVA with SOA as a factor. Moreover, we compared the RT of each SOA condition with that of the visual-only condition using paired t-tests. Multiple comparisons were corrected through the false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2001).

EEG Recording and Preprocessing

Electroencephalogram (EEG) was recorded via a 64-channel MR-compatible system (Brain Products, GmbH, Gilching, Germany). The system included 63 scalp electrodes (sintered Ag/AgCl passive electrodes) and an additional electrocardiogram (ECG) electrode was attached to the back of participants to control for cardioballistic artifacts. The scalp electrodes were mounted on an elastic cap (BrainCap MR, Brain Products, GmbH) according to the extended 10/20 system. The FCz and AFz scalp electrodes were used as the reference and ground electrodes, respectively. No further offline re-referencing was applied. Impedances at all recording electrodes were typically set below 10 kΩ by applying conductive paste (ABRALYT

2000, FMS, Herrsching–Breitbrunn, Germany). EEG signals were acquired at a 5-kHz sampling rate and band-pass-filtered between 0.016 and 250 Hz.

EEG data were analyzed offline using Brain Vision Analyzer 2.0 (Brain Products, GmbH), the Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011), and our custom MATLAB scripts (The MathWorks). EEG preprocessing steps were similar to those described previously (Kaya et al., 2017). First, the data were down-sampled to 500 Hz and the cardioballistic artifacts were removed by the signal from the ECG channel (Allen, Polizzi, Krakow, Fish, & Lemieux, 1998). Second, the data were filtered through a zero-phase shift Butterworth high-pass filter (3 Hz, 24 dB/octave) and a 50-Hz notch filter (50 Hz ± 2.5 Hz, 16^{th} order). Previous research indicated that different levels of expectancy can originate in dynamic modulation of the delta oscillation phase (1-3 Hz). The low-frequency oscillations in this range play a functional role in human anticipatory mechanisms (Stefanics et al., 2010). It was also shown that slow oscillatory activity (1-3 Hz) related to intersensory attention may entrain to regular stimulation and hence affect the evoked activities (Gomez-Ramirez et al., 2011). Similar to previous multisensory studies (e.g., Keil, Pomper, Feuerbach, & Senkowski, 2017), we used a 3 Hz cut-off frequency for high-pass filtering to limit the contribution of this possible confound. We also confirmed that this filtering procedure did not introduce a significant artifact in the final identified electrode locations and time window. For bimodal and visual-only conditions, the event marker was set at the onset of the visual bar and this time point was considered as the reference zero-point in time. For the auditory-only condition, the reference point was adjusted to the onset of click (which corresponded to the onset of the visual bar in the timeline of bimodal and visual-only conditions). Then, the data were segmented into epochs from -600 ms to 1000 ms. At the final stage, the infomax independent component analysis was applied to these epochs to remove common EEG artifacts such as eye blinks. The components were evaluated according to each participant's scalp maps and activity profiles (Jung et al., 2000). Around 3 components (M = 2.65, SD = 1.87) were typically removed. Each trial was screened automatically by artifact rejection criteria and manually by eye. In the automatic artifact rejection, any trial with oscillations over 50 μV/ms or a voltage change of more than 200 μV was rejected. Any missing and excessive noisy channels (M = 1.16, SD = 1.95) were interpolated using a spherical-spline procedure (Perrin, Pernier,

Bertrand, & Echallier, 1989). Trials with artifacts (on average 11.77% of trials per condition, SEM = 2.47%) were rejected from further ERP and behavioral data analyses.

ERP Analyses

After the preprocessing steps, EEG signals from each specific electrode were averaged across trials to compute ERPs and a low-pass filter (6^{th} order zero-phase Butterworth IIR filter with 40 Hz cut-off frequency) was applied to further smooth these ERPs. Baseline correction was applied according to the -260 to -160 ms before the onset of the visual bar (and the corresponding time point in the auditory-only condition). For all the conditions, this time range was before the onset of the first stimulus and there was no stimulation. In the experimental paradigm studied here, observers performed a speeded discrimination task on the location of visual flash while listening to the static click passively. In other words, vision and audition were primary task-relevant and secondary task-irrelevant modalities, respectively. As in previous studies, we expected to find significant effects of auditory timing on visual reaction times. This pattern of results would imply that the information provided by audition interacts and interferes with the processing primarily carried out by vision. Accordingly, our ERP analyses were based on an application of the additive model [(AV-A) vs. V or AV vs. (A+V)] to detect nonlinear neural response interactions and to reveal modulations of these nonlinear components by auditory timing (see Stevenson et al., 2014, for a review and comparison of models). This approach has been commonly used in EEG studies on humans to quantify audiovisual interactions (e.g., Cappe, Thut, Romei, & Murray, 2010; Giard & Peronnet, 1999; Molholm et al., 2002; Raij et al., 2010). More importantly, the application of this model to ERPs revealed a similar timeline of audiovisual interactions to that of analysis employing reference-independent global measures of the electric field at the scalp (Cappe et al., 2010).

To identify SOA dependent modulations of nonlinear neural response interactions, we first subtracted the auditory-only ERPs from those elicited by bimodal stimulation (AV-A). For each participant and electrode location, the auditory-only epoch (i.e., ERP without baseline correction) was first extracted and aligned to match stimulus onset according to the SOA used in the bimodal condition (AV). Then, this waveform was baseline corrected using the same pre-stimulus time range as the one used for bimodal conditions (-260 to -160 ms). To quantify non-linear

audiovisual interactions, this synthetic ERP was subtracted from the corresponding AV condition. Hence, the difference (AV-A) ERP for each SOA condition was computed. To determine the spatiotemporal profile of significant modulations by auditory timing, we performed running repeated-measures ANOVAs (with SOA as a factor) on the difference (AV-A) ERPs for each time point and electrode location. It should be noted that an ANOVA (or a correlation) 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 eight SOA conditions in the latter one. To overcome multiple comparisons across time and electrode location at the cluster-level, we used the cluster-based permutation test integrated into the Fieldtrip toolbox (Maris & Oostenveld, 2007). Briefly, this approach clusters spatially and temporally adjacent samples with *F* values exceeding an uncorrected alpha level of 0.05. We additionally required at least three neighboring electrodes to form a cluster. Then, the cluster-level statistic was calculated by taking the sum of *F* values within a spatiotemporal cluster. Also, a null-distribution of cluster-level statistics was created by using Monte Carlo simulations with 5,000 permutations, in which condition labels were randomly exchanged within each participant. Finally, the observed (i.e., empirical) cluster-level statistics were compared to the generated null-distribution. The observed cluster-level statistics which fell in the highest or the lowest 2.5th percentile of the generated null-distribution were considered to be significant.

In our study, we specifically aimed to reveal auditory modulations that parallel changes in discrimination performance. As detailed above, the main behavioral measure was reaction time (see *Behavioral Data Analysis*). Therefore, the correlations of changes in the difference ERPs with the corresponding mean reaction times were examined at each time point and electrode location. For each SOA condition, the difference ERPs were averaged across participants and their amplitudes were compared with the corresponding RTs, which were also averaged across participants. The relationship between these two measures across different SOA conditions was assessed through linear regression linear fits. As in running ANOVAs, we had calculations of multiple correlations across time and electrode locations. Similar to previous studies (e.g., Colosio, Shestakova, Nikulin, Blagovechtchenski, & Klucharev, 2017; Han, Yoo, Seo, Na, & Seong, 2013; Riberio & Castelo-Branco, 2019), we applied a cluster-based permutation test to solve this problem and to cluster selected samples ($p < 0.05$) objectively. The correlation

coefficients were used to have cluster-level statistics. Other conventions and parameters of the permutation test were the same as those used for the running ANOVAs described above.

Of note, any confounding factor that existed in all the bimodal conditions (i.e., in all the difference ERPs), did not change with auditory timing, and did not correlate with RT value changes were not reported as significant. In other words, any criteria taking both the outcome of the ANOVA and correlation tests into account are expected to be resistant to any confounding factor such as common anticipatory processes that might lead to spurious audiovisual interactions (Besle, Fort, & Giard, 2004; Teder-Sälejärvi, McDonald, Di Russo, & Hillyard, 2002). Therefore, based on the outcome of the ANOVA test and the correlation maps (i.e., significant spatiotemporal clusters), we identified time windows and electrode locations associated with both significant effects of SOA and correlations. We used the identified electrode locations (i.e., exemplar sites) to display evoked brain activity time-courses for illustrative purposes and also performed additional post-hoc tests over these electrode locations. For the identified time window, we computed the mean difference (AV-A) ERP amplitude and tested whether these values are significantly different than that of visual-only (V) baseline level for each SOA value through paired t-tests. Any significant positive or negative deviation was interpreted as a super-additive [AV > (A + V)] or a sub-additive [AV < (A + V)] interaction. Multiple comparisons were corrected using the FDR procedure. Moreover, to further elucidate the source of audiovisual interactions, we computed the peak latencies and amplitudes of the components over the identified electrode locations. Using the specific time range of each component, we computed these metrics for each condition and observer. We performed one-way repeated-measures ANOVA (with SOA as a factor) on these metrics and also carried out a correlation analysis between the modulations of each metric and changes in behavioral reaction time measures by auditory timing. The correlation between these measures across different SOA conditions was also evaluated through linear regression fits having intercept and slope as coefficients.

Results

Behavioral Results

All observers reported the location of the flashed bar with high accuracy (M = 95.69%, SEM = 0.83%), suggesting that they could easily perform the task at near-ceiling levels. There was no

effect of SOA on the percent correct values of AV conditions and none of these percentage values was significantly different than that of V (visual-only) condition. We only used the trials with correct responses in the subsequent estimation of RT values and ERP analyses. Figure 2 shows the average RT values of AV and V conditions. A one-way repeated-measures ANOVA on the RT values of AV conditions revealed a significant effect of SOA ($F_{7,133} = 50.626$, $p < 0.001$, $\eta_p^2 = 0.727$). An increase in the SOA led to an increase in the RT values such that the RTs of negative SOA (i.e., auditory-leading) conditions were smaller than those of positive SOA (visual-leading) conditions. These results suggest that the observers perceived the visual flash and its location earlier in the small negative SOA conditions, and thus leading to smaller RT values when compared to that of positive SOA conditions. Except for 80 and 120 ms SOA, RTs of all other conditions were significantly smaller than that of visual-only (FDR corrected pairwise comparisons, $p < 0.05$). None of the AV conditions was significantly higher than V in terms of RT values.

[FIGURE 2 ABOUT HERE]

Audiovisual Interactions: Time-courses and Scalp Topographies

We performed running repeated-measures ANOVA with cluster-based permutation test on the difference (AV-A) ERPs. Figure 3a displays the outcome of this test. We found two spatiotemporal clusters associated with the significant effect of SOA. The early cluster was within 126-176 ms time range and mainly over medial parietal scalp sites (cluster-level $F_{sum} = 1182.5$, $p = 0.018$). These modulations were also extended over occipital and central electrodes (Figure 3a, c). The later cluster (cluster-level $F_{sum} = 8995.5$, $p < 0.001$) started around 230 ms and these modulations became dominant over almost all electrodes around 300 ms (exact time range: 228-348 ms). As shown by the outcome of additional correlation analysis (Figure 3b), only the early modulations were correlated with the changes in RT values at the cluster-level (120-184 ms time range; cluster-level $t\text{-}stat_{sum} = 3173.8$, $p < 0.001$). For this time range, the correlations were present over medial parietal, centro-parietal and parieto-occipital electrodes. Similar to the outcome of the ANOVA test, these observed correlations were also spread over central and occipital scalp sites.

[FIGURE 3 ABOUT HERE]

Averaged ERP Amplitudes from Exemplar Sites

The electrodes, which were part of early spatiotemporal clusters revealed by both the ANOVA and correlation tests, were selected as exemplar sites. The averaged potentials are shown in Figure 4. Over these electrodes, there were robust evoked activities to the visual flash and auditory click. However, the activities elicited by the click were earlier and had relatively smaller amplitudes. Within the 126-176 ms time range (late P1 and early N1 component range), the scalp topography for the auditory click was also different and the activations were centered over temporal sites (Figure 4a). Simultaneous presentation (SOA = 0 ms) of the visual flash and auditory click overall elicited components with larger amplitudes.

[FIGURE 4 ABOUT HERE]

The averaged difference (AV-A) ERPs for all the SOA conditions are displayed in Figure 4b. Within the 126-176 ms time range, the averaged values for the positive SOAs were significantly higher than those for the negative SOAs (Figure 4c) and they increased when there was an increase in the SOA value. We further compared the averaged ERP amplitude of each SOA condition (i.e., AV-A of each SOA condition) with that of the V baseline level. The averaged values of all the negative SOAs were significantly smaller than the baseline level (FDR corrected pairwise comparisons, $p < 0.05$), suggesting robust sub-additive interactions [AV < (A+V)] for these SOA values. Although the averaged values of positive SOAs were slightly above this level, none of them were significantly different. Another important point is that the changes in the averaged difference ERPs mostly occurred when the absolute value of SOA was smaller than 100 ms. This was consistent with the modulations of behavioral RT values. In other words, both behavioral RT and averaged neural activities (Figure 2, 4c) pointed to a similar morphology of SOA dependency which was supported by running ANOVAs and correlations in the cluster-based permutation test. For these cluster of electrodes centered over medial parietal electrodes and extending over occipital and central sites, the results suggested a robust correlation between RT values and the modulations of ERP components within 126-176 time range (Figure 4d).

To further understand the nature of observed SOA modulations and audiovisual interactions, we additionally performed ANOVA and correlation tests on the peak latencies and amplitudes of P1 and N1 components (Figure 5). These analyses overall pointed to the significant changes in the N1 component rather than P1. In particular, the (peak) amplitude of the N1 component was significantly dependent on SOA and correlated with the changes in RT values (Figure 5b, Table 1). These negative values increased (i.e., the absolute value of amplitude decreased) as the SOA was increased. Moreover, this dependency on SOA and a monotonic linear increase were similar to the one displayed in Figure 4c. There were sub-additive interactions in the negative SOA range corresponding to the enhancement of N1 amplitude (Stekelenburg & Vroomen, 2005). The correlation tests reported significant correlations for the P1 amplitude and N1 latency as well. However, these changes were not significantly dependent on SOA and not meaningful when the whole SOA range was considered. The outcome of these additional tests on each ERP component suggests that the significant changes in the N1 amplitude rather than latency shifts mainly contributed to the observed SOA modulations and audiovisual interactions over the identified medial parietal electrodes.

[FIGURE 5 ABOUT HERE]

[TABLE 1 ABOUT HERE]

Discussion

Using a wide range of SOA values, we investigated audiovisual interactions within the context of a speeded discrimination task on visual flash. The audiovisual interactions, which were within 126-176 ms time range (i.e., within the P1 and N1 components range) and centered over medial parieto-occipital and parietal sites, were modulated by SOA. More importantly, these ERP modulations were also correlated with the changes in RT values. Follow-up analyses revealed that these observed SOA modulations were mainly due to amplitude changes in the N1 component. Within the context of a simple discrimination task, these results highlight the importance of low-level audiovisual interactions within a distinct time window. In particular, these results reveal an important relationship with the final RT values and early ERP components which were not explicitly provided by previous studies focused on event-related oscillations (e.g.,

Naue et al., 2011; Thorne et al., 2011). They also suggest a significant correlation between these modulations and perceived visual timing in multisensory profiles. In the following sub-sections, we discuss the implications of these findings for audiovisual interactions in the temporal domain and for the effects of auditory timing on vision.

Stimulus Asynchrony Effects on Audiovisual Interactions

In the previous phase-resetting studies, either negative (auditory-leading) or positive (visual-leading) SOA values were used based on the primary modality. Using a high sampling rate of SOAs, the main focus of these studies was to indicate a functional link between the modulations (i.e., fluctuations/ripples) of the low-frequency phase coherency values and the final behavioral performance of individual subjects (e.g., Naue et al., 2011; Thorne et al., 2011). Since these studies were mostly restricted to either negative or positive SOA values, they failed to provide a direct relationship between the RT values and modulations of ERPs within a wide range of SOAs. Our findings fill this important gap in the literature and complement these studies. In both RT and ERP metrics (Figure 4c-d), we found a robust monotonic increase in the short SOA range (i.e., -100 ms $<$ SOA $<$ 100 ms). This transition can only be revealed by including both negative and positive SOA range. Due to our relatively low sampling rate of SOAs and data analysis approach (i.e., analysis on the signals averaged across trials), our findings here do not provide direct supporting evidence for the phase-resetting hypothesis. However, in general, they are consistent with the phase-resetting studies by revealing audiovisual interactions in the temporal domain over parieto-occipital scalp sites. Previous phase-resetting studies emphasize strong influences of a preceding secondary stimulus (e.g., a click) on the primary target (e.g., visual flash) and indicated significant audiovisual interactions over the visual cortex (e.g., Naue et al., 2011). This corresponds to our negative (auditory-leading) SOA conditions. We observed significant deviations and decrease from the baseline level for both RT and ERP values mainly in the negative SOA range. Particularly, our findings are in line with these studies by highlighting the importance of negative SOA conditions. An exception is the RT value at +40 ms of SOA. Compared to vision, audition has better temporal resolution and less processing latencies (Burr, Banks, & Morrone, 2009; Spence & Squire, 2003; Rammsayer, Borter, & Troche, 2015; Vroomen & Keetels, 2010). As also indicated by Figure 4a, the evoked activities to auditory stimulation were earlier. Accordingly, in terms of sensory and perceptual processing, a 40 ms

positive SOA may correspond to synchronous stimulation (or might even be in the negative range) in our setting.

Compared to the 126-176 ms time range (late P1 and early N1 component range), previous research has also pointed out audiovisual interactions over earlier or later ERP components. In these studies, the experimental design was mostly restricted to simultaneous (SOA=0) presentation or included only a few SOA conditions (e.g., Mercier et al., 2013; Molholm et al., 2002). Each bimodal difference ERP was compared to the baseline level [i.e., V level for (AV-A) waveforms] to reveal interactions at specific conditions. Based on the cluster-level statistics, our results did not indicate audiovisual interactions over early components associated with the significant effect of SOA. We found SOA dependent modulations over later (around 300 ms) components. However, these modulations were not correlated with the changes in the final behavioral RT values. Moreover, they were present in almost all electrode locations and fluctuated across SOA conditions. In other words, these SOA effects were not meaningful. Although our experimental design and ANOVA tests on the difference ERPs are expected to be resistant to spurious audiovisual interactions, it is still possible that these modulations in difference ERPs may originate from a late common activity present in both unimodal and bimodal conditions (Besle et al., 2004). In terms of scalp topographies, the audiovisual interactions in the 126-176 ms time range were meaningful. The sub-additive effects in this time range have been mainly interpreted as the direct influence of auditory inputs on the sensory processing in the visual cortex (Molholm et al., 2002; Teder-Sälejärvi, Di Russo, McDonald, & Hillyard, 2005; Teder-Sälejärvi et al., 2002). Given that the sub-additive interactions were mainly observed in our negative SOA conditions (i.e., auditory-leading conditions), this interpretation is in line with the current findings. Such direct influence of a preceding click and crosstalk may be achieved through sparse neuroanatomical connections between auditory and visual cortices (Cappe & Barone, 2005; Clavagnier, Falchier, & Kennedy, 2004; Falchier, Clavagnier, Barone, & Kennedy, 2002). Using a combination of basic ERP analyses, reference-independent topographic analyses and source estimations with an audiovisual motion paradigm, Cappe et al. (2010) further indicated that the early sub-additive audiovisual interactions reflect not only strength modulations but also the topographic modulations. The source estimations revealed simultaneous early sub-additive effects within a network of primary visual, primary auditory

cortices and posterior superior temporal sulcus. This further points to a more elaborate network and suggests that functional coupling between these regions may underlie these interactions. It is important to note that our findings revealed strength modulations at specific cluster of electrodes rather than major shifts in the scalp topography. Given the flexible and adaptive nature of multisensory processing (van Atteveldt et al., 2014), this may be due to the differences in criterion content (i.e., motion perception) and stimulation profile. We revisit this issue in the following sub-section.

Since we characterized behavioral RT values and ERP measures within a wide range of SOA values, we were able to distinctively observe the effects of SOA rather than the temporal order between two events. For instance, the modulations within the 126-176 ms time window cannot be explained only by a change in the order of events. An account purely based on the temporal order suggests an overall difference between negative and positive SOA values, but this difference should not be modulated by a change in the absolute amount of asynchrony (i.e., step function). However, both ERP and RT modulations did not suddenly change when there was a change in the sign of SOA. In both datasets (Figure 2 and 4c), there was a gradual but robust linear increase within the short SOA range (i.e., -100 ms < SOA < 100 ms). Previous studies have shown that human observers have very low performance in a temporal order judgment task and do not even perceive the order of visual and auditory events in this SOA range (Vroomen & Keetels, 2010). Using the SOA values covering this important range, Talsma, Senkowski, and Woldorff (2009) investigated the effect of intermodal attention on audiovisual interactions in time. In their audiovisual conditions, the participants attended to either auditory or visual stimulation while detecting an occasional target in the attended modality (see also Senkowski Talsma, Grigutsch, Herrmann, & Woldorff, 2007, for a similar experimental design). Their results also highlight the importance of modulations within the P1 and N1 component range. On the other hand, they were not able to show a direct relationship between these modulations and the final response times since there was no significant effect of SOA and/or a two-way interaction between SOA and attention on the measured RT values. Building on these findings, it is expected that attentional cueing and alerting have limited contributions to the identified SOA range in which human observers do not even perceive the order of auditory and visual stimulation. Any attentional cueing and alerting may take place at SOA values longer than 100 ms (e.g., -160 ms).

Previous research also indicated that subcortical areas and non-specific pathways contribute to audiovisual processing (e.g., van den Brink et al., 2014). It is still possible that a preceding click can engage these areas and lead to earlier interactions related to attentional cueing and alerting mechanisms. This possibility cannot be ruled out with neural recordings from the scalp surface. Future systematic investigations will be informative in this respect.

Auditory Timing for Different Aspects of Vision

In the current EEG study, the observers performed a discrimination task rather than a task directly engaging perceived timing. However, previous research has revealed that a decrease in RT value in a speeded discrimination task reflects behavioral facilitation due to enhanced visual processing in bimodal presentation (Dochin & Lindsey, 1966; Molholm et al., 2002). The modulations of RTs have been associated with the behavioral outcome of perceptual tasks engaging perceived timing (Cardoso-Leite, Gorea, & Mamassian, 2007). Of particular interest here, the speeded RTs have been commonly used by previous multisensory studies to quantify perceptual shifts in the temporal domain (Diederich et al., 2012; Navarra et al., 2009). Accordingly, our results also provide important implications for understanding common and distinct processes that take place in different experimental designs on both audiovisual stimulation and perceived visual timing. For example, using an experimental design based on a flash-lag paradigm, Stekelenburg and Vroomen (2005) examined the effects of click timing (i.e., auditory timing) on the perceived timing of a visual flash and the early ERP components elicited by the visual flash. Compared to the synchronous presentation of click, the visual flash was perceived earlier if the click preceded the visual flash. Conversely, a click presented after the flash made the flash perceived later. In addition to these changes in the perceived timing of visual flash, they found significant modulations in the amplitude (but not in the latency) of N1 component over the parieto-occipital scalp sites. More importantly, these modulations were also correlated with perceptual changes. These initial findings are interesting and novel by highlighting the role of low-level audiovisual interactions in the observed perceptual changes. On the other hand, the experimental design was only restricted to the leading (SOA=-100 ms), synchronous (SOA=0) and lagging (SOA=100) conditions. Although the time range of the significant modulations and correlations presented here do not exactly match with the one reported by Stekelenburg and Vroomen (2005), our results based on a rich repertoire of temporal

profiles support their findings. They overall suggest that audiovisual interactions (which were elicited by an auditory and a visual event) in the N1 component play an important role in the effects of auditory timing on perceived visual timing.

The effects of auditory timing on other visual features have been demonstrated by relatively more complex audiovisual stimulations (e.g., Freeman & Driver, 2008; Getzmann, 2007; Morein-Zamir, Soto-Faraco, & Kingstone, 2003). In the motion domain, two consecutive apparent motion frames (e.g., flashes) with a fixed time interval have been typically used. For auditory stimulation, two concurrent auditory events (e.g., clicks) have been used and the time interval between them is systematically changed. The time interval demarcated by these auditory events has been found to modulate motion perception. For instance, Kafaligonul and Stoner (2010) showed that auditory time intervals can change the perceived speed of two-frame apparent motion (see also Ogulmus, Karacaoglu, & Kafaligonul, 2018). The apparent motion with a short auditory time interval was perceived to move faster than the one with a long time interval. These changes have been mainly explained by describing that auditory clicks drive the timing of apparent motion frames (or the time interval between them). Hence, the shortening and lengthening in the perceived time interval between the motion frames have been considered to result in faster and slower motion percepts, respectively. In a recent EEG study, Kaya and Kafaligonul (2019) investigated the cortical processes underlying these effects of auditory timing on perceived speed. In their design, each apparent motion frame (i.e., visual flash) and each click had the same durations as the ones used here. Their results pointed to both early and late modulations of the neural activity over different scalp sites, suggesting that auditory timing may take place at different stages of motion processing. Interestingly, the earliest modulation of neural activity occurred in the N1 component (150-200 ms time range) over medial parietal and parieto-occipital scalp sites. In terms of stimulation, these early modulations roughly corresponded to the presentation of the first apparent motion frame and click. This is highly similar to our results which were found by using a single auditory and a visual event and by engaging subjects in a speeded discrimination task. On the other hand, the later modulations (490-540 ms) over these electrodes were mostly in agreement with the changes in perceived speed. These late modulations were beyond the completion of apparent motion and the time interval demarcated by clicks (i.e., after the presentation of the second frame and clicks). Accordingly, our results here not only

confirm the earliest interaction by Kaya and Kafaligonul (2019) but also suggest that the early modulations of the N1 component over these scalp sites may be due to the interaction between the first auditory and visual events in these relatively complicated experimental designs and tasks. The later modulations may be specific to the processing of visual features and the relative recruitment of different cortical areas (and associated processes) may be based on the perceptual task engaged in.

Mounting evidence suggests that multisensory integration involves cortical areas at different stages of sensory processing. The current notion also highlights the dynamic recruitment of different cortical areas and processes during integration. Early crossmodal interactions at low-level sensory areas have been considered to be an important part of the integration process and interpreted as reflecting the automatic and stimulus-driven nature of multisensory integration (Talsma et al., 2010; van Atteveldt et al., 2014). Notably, previous studies indicated that early audiovisual interactions in primary sensory cortices highly depend on the temporal and spatial characteristics of stimulation (Chen & Vroomen, 2013). Our findings here are consistent with this view by showing the SOA dependency of early audiovisual interactions in the N1 component. On the other hand, when the modulations of N1 component are compared with previous research (e.g., Kaya & Kafaligonul, 2019), the interactions in this component range also depend on the criterion content and can even be directly correlated with final perceptual performance in a simple visual discrimination task. In line with these findings, previous audiovisual studies emphasize the flexible and highly adaptive nature of subadditive interactions (i.e., nonlinear enhancement of N1 amplitude) in this component (e.g., Fort, Delpuech, Pernier, & Giard, 2002; Giard & Peronnet, 1999). From a broader perspective, such flexible and adaptive feature reflects the dynamic recruitment of integrative processes (even at early stages of sensory processing) which may be important for increasing the efficiency of audiovisual integration for a particular perceptual task.

Conclusion

To sum up, using a speeded discrimination task combined with EEG recording, we investigated the relationship between audiovisual interactions in the temporal domain and behavioral reaction times. The averaged neural activities over medial parietal, parieto-occipital and occipital

electrodes within the 126-176 ms time range were significantly modulated by the relative timing between the auditory and visual events. Moreover, these modulations were correlated with the changes in reaction time values and further analyses suggested that they were mainly due to changes in the amplitude of the N1 component. Together with previous research, these findings highlight the importance of the N1 component in audiovisual temporal processing and also provide evidence that the crossmodal interactions at early stages of sensory processing play a critical role in the final behavioral performance.

Acknowledgments

We are grateful to Can Oluk for technical assistance in data collection. The authors would also like to thank Dr. Murat I. Atagun who was at the origin of this project. This project was supported by the Scientific and Technological Research Council of Turkey (grant number 113K547) and the Turkish Academy of Sciences (TUBA-GEBIP Award).

Conflict of Interest

The authors declare no competing financial interests.

Data Accessibility

The dataset and analysis tools of the current study are available from the corresponding author on request. Any access to the dataset will be in accordance with the informed consent signed by the participants.

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. https://doi.org/10.1006/nimg.1998.0361

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B, 57,* 289–300. http://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics, 29,* 1165–1188. https://doi.org/10.1214/aos/1013699998

Besle, J., Fort, A., & Giard, M. (2004). Interest and validity of the additive model in electrophysiological studies of multisensory interactions. *Cognitive Processing, 5,* 189–192. https://doi.org/10.1007/s10339-004-0026-y

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436. http://doi.org/10.1163/156856897X00357

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

Cappe, C., & Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience*, *22*, 2886–2902. https://doi.org/10.1111/j.1460-9568.2005.04462.x

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2010). Auditory-visual multisensory interactions in humans: Timing, topography, directionality, and sources. *Journal of Neuroscience*, *30*, 12572–12580. https://doi.org/10.1523/jneurosci.1099-10.2010

Cardoso-Leite, P., Gorea, A., & Mamassian, P. (2007). Temporal order judgment and simple reaction times: Evidence for a common processing system. *Journal of Vision*, *7*(6), 11. https://doi.org/10.1167/7.6.11

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. *Journal of Neuroscience, 37,* 5274–5287. https://doi.org/10.1523/jneurosci.2926-16.2017

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

Clavagnier, S., Falchier, A., & Kennedy, H. (2004). Long-distance feedback projections to area V1: Implications for multisensory integration, spatial awareness, and visual consciousness. *Cognitive, Affective and Behavioral Neuroscience*, *4*, 117–126. https://doi.org/10.3758/cabn.4.2.117

Colosio, M., Shestakova, A., Nikulin, V. V., Blagovechtchenski, E., & Klucharev, V. (2017). Neural mechanisms of cognitive dissonance (revised): An EEG study. *Journal of Neuroscience, 37*(20), 5074–5083. https://doi.org/10.1523/jneurosci.3209-16.2017

Diederich, A., Schomburg, A., & Colonius, H. (2012). Saccadic reaction times to audiovisual stimuli show effects of oscillatory phase reset. *PLoS ONE, 7*(10), e44910. https://doi.org/10.1371/journal.pone.0044910

Donchin, E., & Lindsley D. B. (1966). Average evoked potentials and reaction times to visual stimuli. *Electroencephalography and Clinical Neurophysiology, 20,* 217–223. https://doi.org/10.1016/0013-4694(66)90086-1

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience, 22,* 5749–5759. https://doi.org/10.1523/jneurosci.22-13-05749.2002

Fort, A., Delpuech, C., Pernier, J., & Giard, M. H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cerebral Cortex, 12,* 1031–1039. https://doi.org/10.1093/cercor/12.10.1031

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

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

Giard, M. H., & Peronnet, F. (1999). Auditory–visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11*, 473–490. https://doi.org/10.1162/089892999563544

Gomez-Ramirez, M., Kelly, S. P., Molholm, S., Sehatpour, P., Schwartz, T. H., & Foxe, J. J. (2011). Oscillatory sensory selection mechanisms during intersensory attention to rhythmic

auditory and visual inputs: A human electrocorticographic investigation. *Journal of Neuroscience*, *31*, 18556–18567. https://doi.org/10.1523/jneurosci.2164-11.2011

Han, C. E., Yoo, S.W., Seo, S. W., Na, D. L., & Seong, J.-K. (2013). Cluster-based statistics for brain connectivity in correlation with behavioral measures. *PLoS One, 8,* e72332. https://doi.org/10.1371/journal.pone.0072332

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology*, *111*, 1745-1758. https://doi.org/10.1016/s1388-2457(00)00386-2

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

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

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

Keil, J., Pomper, U., Feuerbach, N., & Senkowski, D. (2017). Temporal orienting precedes intersensory attention and has opposing effects on early evoked brain activity. *NeuroImage*, *148*, 230–239. https://doi.org/10.1016/j.neuroimage.2017.01.039

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1)*,* 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mercier, M. R., Foxe, J. J., Fiebelkorn, I. C., Butler, J. S., Schwartz, T. H. & Molholm, S. (2013). Auditory-driven phase reset in visual cortex: Human electrocorticography reveals mechanisms of early multisensory integration. *NeuroImage, 79,* 19–29. https://doi.org/10.1016/j.neuroimage.2013.04.060

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. *Cognitive Brain Research, 14,* 115–128. https://doi.org/10.1016/s0926-6410(02)00066-6

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research, 17,* 154–163. https://doi.org/10.1016/s0926-6410(03)00089-2

Murray, M. M., Lewkowicz, D. J., Amedi, A., & Wallace, M. T. (2016). Multisensory processes: A balancing act across the lifespan. *Trends in Neurosciences, 39,* 567–579. https://doi.org/10.1016/j.tins.2016.05.003

Murray, M. M., & Wallace, M. T. (2012). *The neural bases of multisensory processes*. Boca Raton, FL: CRC Press.

Naue, N., Rach, S., Strüber, D., Huster, R. J., Zaehle, T., Körner, U., & Herrmann, C. S. (2011). Auditory event-related response in visual cortex modulates subsequent visual responses in humans. *Journal of Neuroscience, 31,* 7729–7736. https://doi.org/10.1523/JNEUROSCI.1076-11.2011

Navarra, J., Hartcher-O'Brien, J., Piazza, E., & Spence, C. (2009). Adaptation to audiovisual asynchrony modulates the speeded detection of sound. *Proceedings of the National Academy of Sciences USA, 106,* 9169–9173. https://doi.org/10.1073/pnas.0810486106

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

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011,* 156869. https://doi.org/10.1155/2011/156869

Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. https://doi.org/10.1163/156856897X00366

780 Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential
781 and current density mapping. *Electroencephalography and Clinical Neurophysiology, 72,*
782 184–187. https://doi.org/10.1016/0013-4694(89)90180-6

783 Raij, T., Ahveninen, J., Lin, F. H., Witzel, T., Jääskeläinen, I. P., Letham, B., … Belliveau, J. W.
784 (2010). Onset timing of cross-sensory activations and multisensory interactions in auditory
785 and visual sensory cortices. *European Journal of Neuroscience*, *31*, 1772–1782.
786 https://doi.org/10.1111/j.1460-9568.2010.07213.x

787 Rammsayer, T. H., Borter, N., & Troche, S. J. (2015). Visual-auditory differences in duration
788 discrimination of intervals in the subsecond and second range. *Frontiers in Psychology*, *6*,
789 1626. https://doi.org/10.3389/fpsyg.2015.01626

790 Ribeiro, M. J., & Castelo-Branco, M. (2019). Neural correlates of anticipatory cardiac
791 deceleration and its association with the speed of perceptual decision-making, in young and
792 older adults. *NeuroImage, 199,* 521-533. https://doi.org/10.1016/j.neuroimage.2019.06.004

793 Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S., & Woldorff, M. G. (2007). Good
794 times for multisensory integration: Effects of the precision of temporal synchrony as revealed
795 by gamma-band oscillations. *Neuropsychologia, 45,* 561–571.
796 https://doi.org/10.1016/j.neuropsychologia.2006.01.013

797 Spence, C., & Squire, S. B. (2003). Multisensory integration: Maintaining the perception of
798 synchrony. *Current Biology*, *13*, R519–R521. https://doi.org/10.1016/s0960-9822(03)00445-7

799 Stefanics, G., Hangya, B., Hernadi, I., Winkler, I., Lakatos, P., & Ulbert, I. (2010). Phase
800 entrainment of human delta oscillations can mediate the effects of expectation on reaction
801 speed. *Journal of Neuroscience*, *30,* 13578–13585. https://doi.org/10.1523/jneurosci.0703-
802 10.2010

803 Stekelenburg, J. J., & Vroomen, J. (2005). An event-related potential investigation of the time-
804 course of temporal ventriloquism. *NeuroReport, 16,* 641–644.
805 https://doi.org/10.1097/00001756-200504250-00025

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 Topography, 27,* 707 –730. https://doi.org/10.1007/s10548-014-0365-7

Spence, C. (2018). Multisensory perception. In I. J. Wixted (Ed.), *Stevens' handbook of experimental psychology and cognitive neuroscience* (Vol. 2, pp. 1-56). Hoboken, NJ, USA: John Wiley & Sons. https://doi.org/10.1002/9781119170174.epcn214

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

Talsma, D., Senkowski, D., Woldorff, M. G. (2009). Intermodal attention affects the processing of the temporal alignment of audiovisual stimuli. *Experimental Brain Research, 198,* 313–328. https://doi.org/10.1007/s00221-009-1858-6

Teder-Sälejärvi, W. A., Di Russo, F., McDonald, J. J., & Hillyard, S. A. (2005). Effects of spatial congruity on audio-visual multimodal integration. *Journal of Cognitive Neuroscience*, *17*(9), 1396–1409. https://doi.org/10.1162/0898929054985383

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. *Cognitive Brain Research, 14,* 106–114. https://doi.org/10.1016/s0926-6410(02)00065-4

Thorne, J. D., & Debener, S. (2014). Look now and hear what's coming: on the functional role of cross-modal phase reset. *Hearing Research, 307,* 144–152. https://doi.org/10.1016/j.heares.2013.07.002

Thorne, J. D., De Vos, M., Viola, F. C., & Debener, S. (2011). Cross-modal phase reset predicts auditory task performance in humans. *Journal of Neuroscience, 31,* 3853–3861. https://doi.org/10.1523/jneurosci.6176-10.2011

van Atteveldt, N., Murray, M. M., Thut, G., & Schroeder, C. E. (2014). Multisensory integration: Flexible use of general operations. *Neuron, 81,* 1240–1253. https://doi.org/10.1016/j.neuron.2014.02.044

van den Brink, R. L., Cohen, M. X., van der Burg, E., Talsma, D., Vissers, M. E., & Slagter, H. A. (2014). Subcortical, modality-specific pathways contribute to multisensory processing in humans. *Cerebral Cortex*, *24*(8), 2169–2177. https://doi.org/10.1093/cercor/bht069

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

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

Zhou, H. Y., Cheung, E. F. C., & Chan, R. C. K. (2020). Audiovisual temporal integration: Cognitive processing, neural mechanisms, developmental trajectory and potential interventions. *Neuropsychologia, 140*, 107396. https://doi.org/10.1016/j.neuropsychologia.2020.107396

Figure Legends

Figure 1. Experimental design. (a) The visual stimulus was a flashed bar either at the left or right of the red fixation point. The timeline for the visual-only condition is displayed at the top and the black filled rectangle in the timeline corresponds to the flashed bar. (b) The auditory stimulus was a brief static click introduced binaurally through earphones. The timeline for the auditory-only condition is displayed at the top and the open (unfilled) rectangle in the timeline corresponds to the click. (c) The timeline for bimodal (AV) conditions. Eight SOA conditions were used and the timeline for each SOA is displayed in separate rows. Relative durations of visual and auditory events are indicated by the thickness of rectangles.

Figure 2. Behavioral Results (n=20). Reaction time values of bimodal conditions as a function of SOA. Error bars indicate standard error (± SEM) across participants. The dotted line indicates the mean value for the visual-only condition and the error bars placed over the symbol on the right represent standard error. A significant difference between each time interval condition and the

visual-only condition was marked with an asterisk sign (FDR corrected two-tailed paired t-test, $p < 0.05$).

Figure 3. Time courses and scalp topographies. (a) Running repeated-measures ANOVAs with the cluster-based permutation test on the difference (AV-A) waveforms. Time is displayed on the abscissa from 0 to 350 ms (relative to the onset of visual flash), and electrodes are displayed on the ordinate. A data point was shaded when there was a significant effect of SOA (uncorrected alpha criterion $p < 0.05$). The significant and nonsignificant spatiotemporal clusters were shaded by black and gray, respectively. Voltage topographical map of the averaged *F* values within the time range of early cluster is displayed at the bottom. The uncorrected significance level is also marked on the color bar. The electrodes, which were part of the significant spatiotemporal cluster for at least 20 ms of contiguous data in the time window, are marked by filled circles on the topographical map. (b) Running correlation analyses with the cluster-based permutation test on the difference (AV-A) waveforms. Voltage topographical map of the averaged *t* values (derived from correlation coefficients) within the time range of significant cluster is displayed at the top. Other conventions are the same as those in the upper plot. (c) Voltage topographical maps of the averaged difference [left: AV-A, right: AV-(A+V)] waveforms (i.e., difference maps) within the identified time window (126-176 ms). The difference maps for each SOA condition are shown in separate rows. The voltage topographical map of V (visual-only) condition and the identified electrodes (which were part of both early clusters) are displayed at the top of the left and right column, respectively.

Figure 4. Averaged activities from the exemplar scalp sites (n=20). (a) Grand-averaged ERPs for the synchronous (SOA=0) condition. The bimodal, unimodal, and derived waveforms are shown with different colors. (b) The difference (AV-A) waveforms of all SOA conditions used. (c) Averaged difference waveform amplitudes in the identified time window (126-176 ms) as a function of SOA. Error bars indicate standard error ($\pm$ SEM) across participants. The dotted line indicates the mean value for the baseline level (V condition), and the error bar placed over the symbol at the end of this line represents $\pm$ SEM. A significant deviation from the baseline level for each condition was marked with an asterisk sign (FDR corrected two-tailed paired t-test, $p < 0.05$). (d) Averaged difference waveforms in the identified time window (126-176 ms) with the

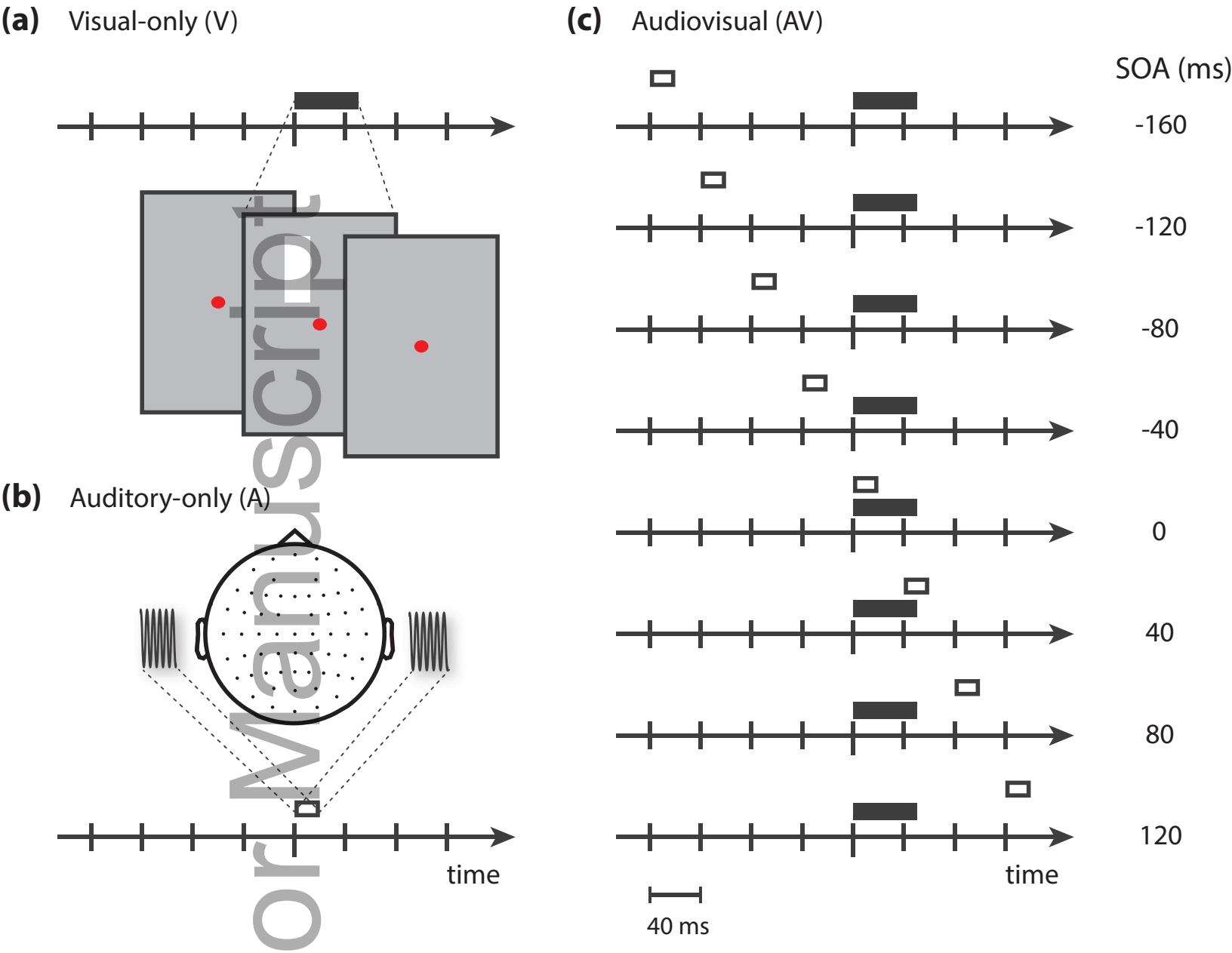
RT values for each SOA condition. Vertical and horizontal error bars correspond to the variance across participants (±SEM). The black solid line indicates the best linear fit and dotted lines denote the 95% confidence intervals (CIs) on the linear fit.

Figure 5. Peak amplitudes and latencies of P1 (a) and N1 (b) components (n=20). The plots on the left display mean values as a function of SOA. On the right, these values are presented with behavioral RTs for each SOA condition. The black solid lines in the right plots indicate the best linear fit and dotted lines denote the 95% CI on the linear fit. Goodness-of-fit of the linear model provided as R^2 along with the corresponding *p* values in Table 1. Other conventions are the same as those in Figure 4c-d.

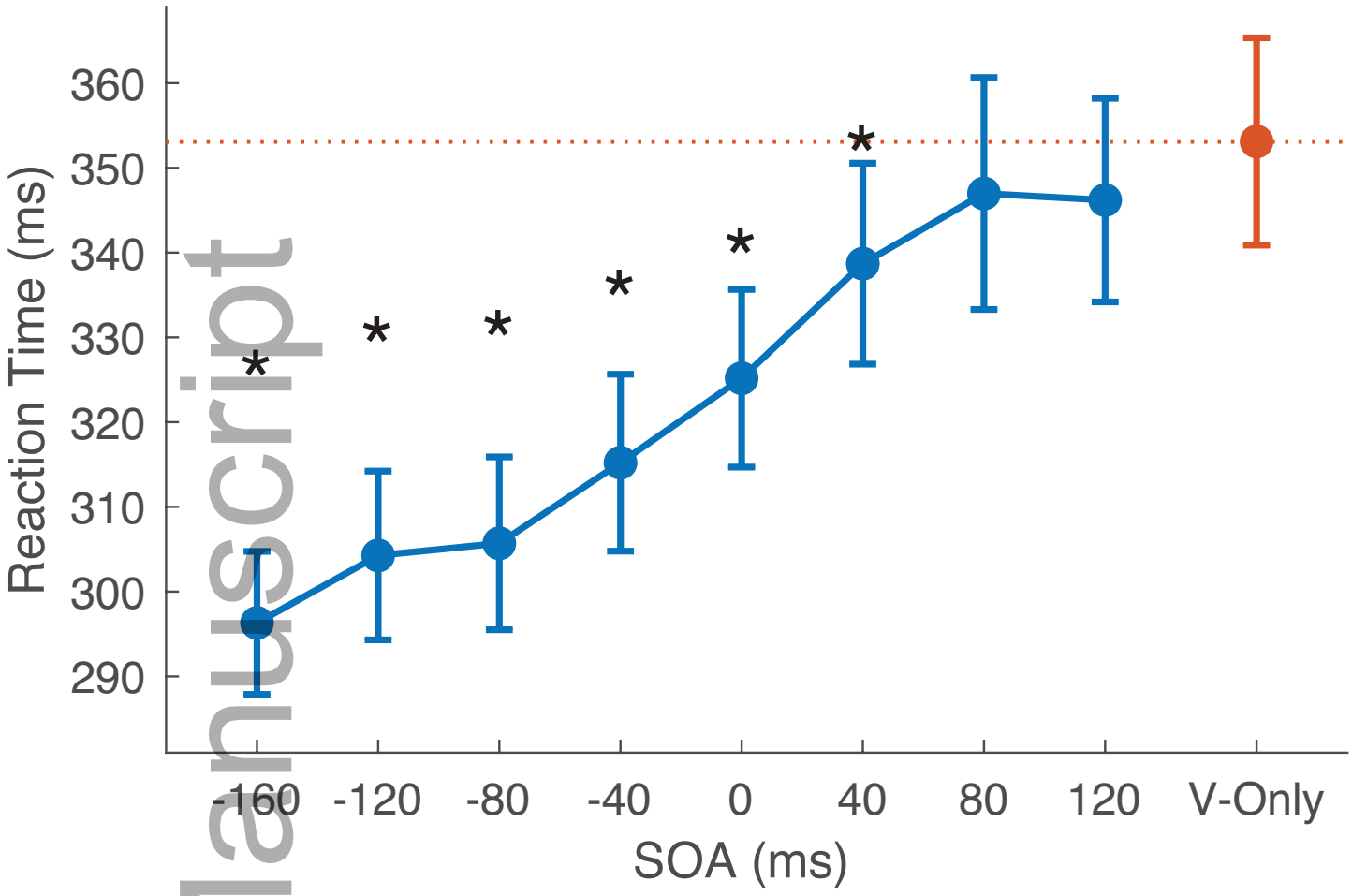
Tables

Table 1. The results of ANOVA and correlation tests on the P1 and N1 components (Figure 5). The values of each component are grouped into separate rows. For each component, the outcome of tests on peak amplitudes are shown first. Significant *p* values ($p < 0.05$) are highlighted in bold.

	ANOVA			Correlation	
	$F_{7,133}$	p	η_p^2	R^2_{adj}	p
P1					
Amplitude	1.530	0.162	0.075	0.526	**0.025**
Latency	0.657	0.708	0.033	0.053	0.282
N1					
Amplitude	2.438	**0.022**	0.114	0.478	**0.035**
Latency	1.376	0.220	0.068	0.433	**0.045**



psyp_13777_f1.eps



psyp_13777_f2.eps