SHORT COMMUNICATION

An ERP Study on Sound-contingent Visual Motion Perception

Ryo TERAOKA[1,*], Osamu WATANABE[2] and Wataru TERAMOTO[3]

[1]*Research Institute of Electrical Communication and Graduate School of Information Sciences,
Tohoku University, Sendai 980-8577, Japan*
[2]*Department of Information Science and Systems Engineering, Muroran Institute of Technology,
Muroran 050-8585, Japan*
[3]*Department of Psychology, Faculty of Letters, Kumamoto University, Kumamoto 860-8555, Japan*

Recent studies have demonstrated that a new association between auditory sequence and visual motion can be rapidly formed even in adults (sound-contingent visual motion perception: SCVM). The present study investigated the neural mechanisms underlying SCVM using event-related potential (ERP). During a 9-min exposure phase, two circles placed side-by-side were alternately and repeatedly presented to produce horizontal apparent motion. The onsets of the circles were synchronized with high- and low-frequency tones. ERPs were measured before and after the exposure phase in response to a pair of visual stimuli (S1 and S2) sequentially presented at a fixed location with a sound sequence that contained the same high and low frequency tones as presented during the exposure phase. After the exposure, the participants experienced SCVM. Significant ERP differences between the pre- and post-exposure were observed not only in the S2 presentation, but also in the S1 presentation. These findings suggest that SCVM processing starts in higher stages of the visual system as early as S1 presentation and then sends feedback signals to earlier stages to induce motion perception at S2 presentation.

KEYWORDS: audiovisual interaction, event-related potential (ERP), associative learning, motion perception, sound-contingent visual motion perception

1. Introduction

Integrating information from different modalities is important for a robust and stable perceptual world, but how does the brain know which signals in different sensory modalities originate from a common external event? Spatiotemporal consistency [1] and temporal correlation [2] of signals provide cues for this task. However, because each sensory modality receives many inputs each second, it seems implausible that the brain depends only on spatiotemporal consistency. A strategy in which the brain forms associations between signals would likely help the brain bind signals efficiently. Recent studies have shown that the brain rapidly forms an association between a sound sequence and visual motion when concurrently presented. Once this association has been formed, the sound sequence can be a driver for visual motion perception. This phenomenon is known as sound-contingent visual motion perception (SCVM; [3]). In these studies, two circles placed side-by-side were alternately presented in conjunction with the alternation of high- and low-frequency tones for several minutes during an exposure phase. Following this exposure phase, participants reported that the presented tones induced illusory motion perception in a static visual stimulus. Considering that this type of association effect occurs in other combinations of sensory modalities, such as between vision and touch [4] and between olfactory and vision [5], forming associations may be a general mechanism across sensory modalities.

Interestingly, SCVM was specifically observed for the previously exposed visual field [3] and eye [6]. These behavioral findings suggest that the associations are formed relatively early visual processing. The aim of the present study was to investigate event-related potentials (ERPs) before and after the formation of an association between a sound sequence and visual motion to clarify the neural mechanisms underlying SCVM.

2. Materials and Methods

Forty-two male volunteers participated in this experiment (age range: 18–25 years, mean: 21.4 years). All participants were naïve to the purpose of the experiment, had normal or corrected-to-normal vision, and had normal hearing. Informed written consent was obtained from all participants before they underwent the experimental procedures, which were approved by the local ethics committee of the Muroran Institute of Technology.[1]

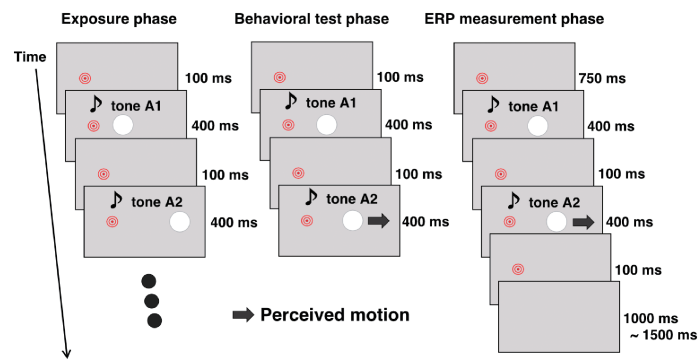[1]This study was conducted in Muroran Institute of Technology.

Fig. 1.   Experimental design. During the exposure phase, participants were exposed to visual apparent motion for 9 min, in which two white circles placed side-by-side were alternately presented. The onset of the leftward circle was accompanied by a high-frequency tone (2 kHz) and that of the rightward circle was accompanied by a low-frequency tone (500 Hz). During the event-related potential (ERP) measurement phase, a white circle was presented twice 10° to the right of the fixation point (red circles). During the behavioral phase, two white circles were presented with a variety of displacements in the left or right direction to measure the magnitude of perceived motion (motion nulling procedure).

Visual stimuli were presented on a 22-inch CRT display (60 Hz) at a viewing distance of 62 cm. White circles (2° in diameter, 5.2 cd/m^2) were presented as visual stimuli on a black background. Three red concentric rings (diameters: 0.143°, 0.429°, and 0.716°; 15.5 cd/m^2) were also presented 10° to the left of the center of the display for fixation point. Auditory stimuli were tone bursts (sampling frequency: 44.1 kHz; 50 ms included a 5-ms rise and fall time; 86 dB A-weighted SPL when presented continuously) delivered to both ears through inner-ear headphones. Participants were instructed to place their heads on a chin rest during the phases. All experiments were conducted in a dark room.

The experiment consisted of three phases: pre-exposure, exposure, and post-exposure. During the exposure phase, two white circles were alternately presented at 7.5° and 12.5° to the right of fixation point to produce apparent horizontal motion. The stimulus-onset asynchrony (SOA) was 500 ms, with a duration of 400 ms and an inter-stimulus interval (ISI) of 100 ms. Onsets of the right and left circles were synchronized with a tone burst of high (A1: 2 kHz) and low (A2: 500 Hz) frequencies, respectively. Participants were instructed to focus on the fixation point during the exposure. The total length of the exposure phase was 9 min; participants were allowed a short break every 3 min.

Each pre-exposure and post-exposure phase consisted of four phases in which the magnitude of illusory motion perception and ERPs were independently and alternately measured twice in this order (perception and ERP measurement phases, respectively). In the behavioral phases, the point of subjective stationarity (PSS) was measured using the method of constant stimuli. The white circle was presented twice (S1 and S2) (SOA: 500 ms, ISI: 100 ms) while being displaced ±0.12°, ±0.24°, ±0.48°, or ±0.96°. Positive and negative values indicate rightward and leftward displacements, respectively. The centers of the two circles were 10° from the fixation point. The onsets were synchronized with those of the same tone bursts (A1 and A2) as in the exposure phase. The frequency of the tones alternated from 500 Hz (A2) to 2 kHz (A1) or from 2 kHz (A1) to 500 Hz (A2). The former and latter frequency changes were accompanied by rightward and leftward visual motion in the exposure phase, respectively. These sound sequences were called the rightward-LH and leftward-HL sound conditions in the pre- and post-exposure phases. Participants were asked to judge whether the visual stimulus moved leftward or rightward. Each phase consisted of 160 trials: 8 displacement conditions × 2 auditory conditions × 10 repetitions. The presentation of each condition was randomized and counterbalanced. The post-exposure phase began soon after the exposure phase and included 10 seconds of readaptation every 20 trials. In the pre- and post-ERP measurement phases, the white circle was presented twice (S1 and S2) at 10° to the right of the fixation point (SOA: 500 ms, ISI: 100 ms). The onsets were synchronized with those of the same tone bursts (A1 and A2) as in the exposure phase. The rightward-LH and leftward-HL sound conditions were conducted as in the perception measurement phases. Each phase consisted of 100 trials: 2 auditory conditions × 50 repetitions. One minute of readaptation was included at the beginning of each phase to prolong the effect of the exposure. The fixation point disappeared with S2 and reappeared 1 s before the presentation of S1. The duration of fixation point disappearance was 1.35 s. Participants were instructed to focus on the fixation point and to refrain from blinking during the presentations of the fixation point to avoid blink and eye movement artifacts.

ERP data were obtained from 31 active electrodes placed according to an extended international 10–20 system: Fp1, Fp2, Fz, F3, F4, F7, F8, FCz, FC1, FC2, FC5, FC6, Cz, C3, C4, CP1, CP2, CP5, CP6, TP9, TP10, Pz, P3, P4, P7, P8, Oz, O1, O2, PO9, and PO10, with a reference electrode placed on FCz. Eye blinks and vertical eye movements were monitored. Data were collected at a sampling rate of 1 kHz. The impedances of all electrodes were kept below 5 kΩ using the BrainAmp system (Brain Products, Germany). Data were analyzed offline using EEGLAB software [7]. Trials in which the amplitude exceeded ±60 μV were discarded as artifact trials. If the number of trials did not match between the pre- and post-exposure phases, trials were randomly excluded from the phase that contained more trials to maintain consistency in the number of trials between phases for each participant. Ten participants were excluded from data
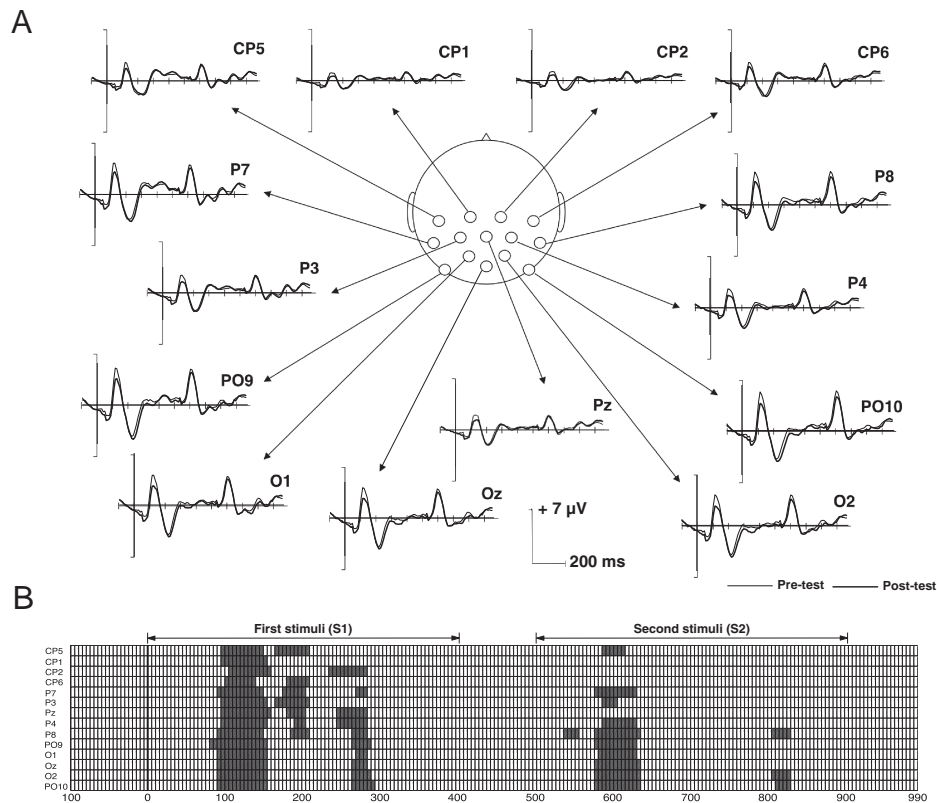
Fig. 2. A: ERP waveforms at central, parietal, and occipital electrodes in the pre- and post-exposure phases. These waveforms were calculated over all trials (except artifact trials) for each test and each electrode position ($n = 22$). The vertical and horizontal axes represent time (ms) and ERP in μV, respectively. Thin and thick lines represent ERP waveforms from pre- and post-exposure phases, respectively. Overall, clearly visible differences appear at all components and all peaks. B: Time intervals in which the amplitude was significantly different between the pre- and post-exposure. Black cells indicate significant difference from zero ($p < 0.05$). While the x-axis indicates time from -100 to 990 ms, the y-axis indicates electrode locations.

analysis because the number of valid trials was less than 75% of the total number in each condition. For valid trials, an epoch was defined from 100 ms before the onset of S1 to 990 ms after the onset of S1. A baseline correction was made using the mean voltage from -100 to 0 ms. Because this study focused on ERPs associated with illusory motion perception, signals from 14 electrodes placed in the occipital and parietal lobes were the focus of the present analysis (CP1, CP2, CP5, CP6, Pz, P3, P4, P7, P8, Oz, O1, O2, PO9, and PO10).

3. Results and Discussion

The PSS value, the 50% point of rightward motion perception, was obtained by fitting a cumulative normal distribution function using a maximum likelihood curve-fitting technique. Ten participants were excluded because their responses were too unstable to fit the function. Data from the remaining 22 participants were analyzed. To investigate whether SCVM occurred, the PSSs of the rightward-LH and leftward-HL sound conditions in each test phase were compared using a two-tailed t-test. While there was no difference in the pre-exposure phase (PSS for the leftward-HL sound condition: 0.049 ± 0.020 [mean $\pm$ standard deviation], PSS for the rightward-LH sound condition: 0.086 ± 0.026; $t_{21} = 1.66$, $p = 0.105$), a significant difference was observed in the post-exposure test (leftward-HL: 0.042 ± 0.017, rightward-LH: 0.079 ± 0.023; $t_{21} = 2.60$, $p = 0.046$). Although the difference was small, these results confirmed that after prolonged, the tones became drivers for the perception of illusory motion.

Figure 2A shows the average waveforms obtained at the 14 targeted electrodes during the pre- and post-exposure phases; the average waveforms were calculated over all accepted trials and all accepted participants. The responses to S1 and S2 both comprised a triphasic waveform with an initial negative component at approximately 75 ms (N75), followed by a positive component at approximately 100 ms (P100), and a negative component at approximately 200 ms (N200). To investigate the ERP signal changes related to SCVM, the amplitude of the difference wave between the pre- and postexposure tests in each sound condition was compared with zero using the t-test for each time sample at each electrode across participants. The time intervals that were significantly different for at least 15 consecutive time samples (15 ms) were considered stable audiovisual interaction intervals [8]. Figure 2B shows significant differences in the S1- and S2-related components between the pre- and post-exposure. The significant differences in S1-related components were observed in intervals between 80 and 155 ms poststimuli (P100) at all channels, between 162 and

209 ms (N200) at the parieto-central channels, and between 235 and 294 ms (P300) at the centro-parietal channels. For S2-related components, the significant differences between the pre- and post-exposure were observed in intervals between 539 and 637 ms poststimuli (N75-2 and P100-2) at parieto-occipital channels and between 808 and 831 ms (P300-2) at occipital channels. Considering that the identical stimuli were presented in the pre- and post-exposure, the identical waveforms should have been observed. However, several ERP components in several channel locations were changed between the pre- and post-exposure. These results would indicate a possibility that the audiovisual processing in several stages is changed after the association between the sound sequence and visual motion was established.

Previous behavioral studies have demonstrated several characteristics of SCVM. Teramoto *et al.* [3] tested $20°$ and $5°$ to the right of the fixation point following exposure to audiovisual stimuli at approximately $10°$ to the right of the fixation point. The results showed no transfer between the visual fields. Similarly, Kobayashi *et al.* [6] investigated whether SCVM occurred via the eye not exposed to the audiovisual stimuli and reported no transfer to the unexposed eye. These findings suggest the involvement of relatively early stages of visual processing in the brain. In the present study, the observation of amplitude modulation at N75 for S2 in the occipitoparietal channels further supports the involvement of relatively early perceptual processing stages in SCVM. The modulation of early ERP components by audiovisual interaction has been documented using several tasks. Giard and Perronet [9] used an audiovisual pattern recognition task and reported that ERPs to audiovisual interaction occurred in the occipital area as early as 40 ms poststimulus. Several studies using double-flash illusion have reported that ERPs appear in the occipital area 103 ms [8] or 30–60 ms [10] after the second sound. Molholm *et al.* [11] also reported modulation of ERP components in the occipitoparietal channels during a simple detection task as early as 46 ms poststimulus. Consistent with these previous studies, the present findings suggest that audiovisual interaction occurs at relatively early stages in the visual system.

Interestingly, the modulation of ERP components started when the S1 was presented. This modulation should have little to do with motion perception itself because motion perception cannot be induced until the S2 is presented. A study reported that ERPs in response to S1 in the apparent motion sequence were almost perfectly matched with those to the isolated presentation of the S1 [12]. In addition, ERPs unique to the apparent motion started appearing around 90 ms after S2 was presented. In contrast, significant modulations were already observed in response to S1 in the present study. This finding is more likely to be associated with audiovisual interactions such as preparation for motion perception rather than with the motion perception itself. Given that either a high-low or low-high tone sequence was presented during the test phases, the brain may predict the perceptual outcome by perceiving the first audiovisual stimuli.

In summary, the present study investigated the neural mechanisms underlying SCVM using ERPs. Results showed the modulation of several ERP components. The modulation of ERP components was different from that to apparent visual motion perception reported in a previous study [12]. These findings suggest that cross-modal motion processing starts in higher stages of the visual system as early as the S1 presentation and then sends feedback signals to earlier stages to induce motion perception at the S2 presentation.

REFERENCES

[1] Calvert, G., Spence, C., and Stein, B. E., The Handbook of Multisensory Processes, MIT Press (2004).

[2] Parise, C. V., Spence, C., and Ernst, M. O., "When correlation implies causation in multisensory integration," *Curr. Biol.*, **22**: 46–49 (2012).

[3] Teramoto, W., Hidaka, S., and Sugita, Y., "Sounds move a static visual object," *PLoS One*, **5**: e12255 (2010).

[4] Teraoka, R., and Teramoto, W., "Touch-contingent visual motion perception: Tactile events drive visual motion perception," *Exp. Brain Res.*, **235**: 903–912 (2017).

[5] Kuang, S., and Zhang, T., "Smelling directions: Olfaction modulates ambiguous visual motion perception," *Sci. Rep.*, **4**: 5796 (2014).

[6] Kobayashi, M., Teramoto, W., Hidaka, S., and Sugita, Y., "Indiscriminable sounds determine the direction of visual motion," *Sci. Rep.*, **2**: 365 (2012).

[7] Delorme, A., and Makeig, S., "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, **134**: 9–21 (2004).

[8] Shams, L., Kamitani, Y., Thompson, S., and Shimojo, S., "Sound alters visual evoked potentials in humans," *NeuroReport*, **12**: 3849–3852 (2001).

[9] Giard, M. H., and Peronnet, F., "Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study," *J. Cogn. Neurosci.*, **11**: 473–490 (1999).

[10] Mishra, J., Martinez, A., Sejnowski, T. J., and Hillyard, S. A., "Early cross-modal interactions in auditory and visual early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion," *J. Neurosci.*, **27**: 4120–4131 (2007).

[11] Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., and Foxe, J. J., "Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study," *Cogn. Brain, Res.*, **14**: 115–128 (2002).

[12] Wibral, M., Bledowski, C., Kohler, A., Singer, W., and Muckli, L., "The timing of feedback to early visual cortex in the perception of long-range apparent motion," *Cereb. Cortex*, **19**: 1567–1582 (2009).