Visual apparent motion modulates temporal processing on audiovisual simultaneity

〇Jinhwan Kwon (Tokyo Institute of Technology), Kenichiro Ogawa (Tokyo Institute of Technology), Taiki Ogata (The University of Tokyo), Yoshihiro Miyake (Tokyo Institute of Technology)

Abstract: Time perception is an important topic in the study of the characteristics of multisensory integration. Given a dynamic environment, the relationship between motion perception and temporal perception is very important, but the relationship has not yet been quantified. We investigated the correlation between motion perception and temporal perception through visual apparent motion and audiovisual temporal order judgment task. As a result, we found new evidence that visual motion perception accelerates temporal perception and improves temporal discrimination in audiovisual processing regardless of prediction. In particular, we suggest that the binding property in motion perception may have an effect on binding with other sensory stimuli in multisensory processing.
Keywords: Motion Perception, Temporal Perception, Audiovisual Simultaneity

1. Introduction

Physical time of any events differs from subjective time on human perception. There are various factors on the difference between the physical and subjective time. For example, physical transmission times through air and neural transmission times are different among senses.

With respect to temporal perception, there has been remarkable progress in simultaneity judgment (SJ) task and temporal order judgment (TOJ) task as a psychophysical study to examine temporal factors in multisensory processes [1]. In particular, the TOJ task is known as a way to measure how human perceive temporal synchrony between two types of senses. In TOJ task, point of subjective simultaneity (PSS) and just noticeable difference (JND) are used as the parameters. The PSS represents a subjective time that the stimuli between the senses are perceived at the same time and make it possible to detect which sensory information was captured early or late. The JND indicates an indicator that determine the temporal resolution in cross-modality [1].

Although many studies find that simultaneity or temporal resolution depends on the various sensory information examined in TOJ task [2, 3] there is little evidence concerning the effect of motion perception in audiovisual integration. Also, many findings have demonstrated that a sound dominates visual events in the temporal dimension [4, 5]. However, it remains unclear how motion information influences audio-visual temporal perception. Given dynamic environment, the motion information is influential factor on temporal perception.

In this study, we focused on the correlation between motion perception and temporal perception. Especially, apparent motion is considered as motion information. Motion can be separated into temporal and spatial elements in which we observe movement, and apparent motion is a phenomenon in which spatiotemporal characteristics of movement are well represented. In particular, it is apparent motion that corresponds to time scale for TOJ task among motion stimuli, i.e. apparent motion is fundamental unit of motion in finite time. Visual apparent motion is an optical phenomenon that makes motion appear by the appropriate spatiotemporal interval even despite two discrete stimuli [6, 7].

The purpose of the present study is to investigate how apparent motion affects temporal perception on audiovisual simultaneity. We examined two types of TOJ task experiment. In experiment 1, we examined whether visual apparent motion has an effect on audiovisual TOJ task. So, participants conducted TOJ task under apparent motion condition and normal condition with single flash on audiovisual simultaneity. However, there remained the influence of not only apparent motion but also specific prediction as higher-order brain function because the interval of two flashes under apparent motion condition was constant. Therefore, in experiment 2, we eliminated the influence of prediction by presenting two visual stimuli, the order of intervals being at random.

2. Methods

2.1 Participants

Twelve participants (11 males and one female, with a mean age of 24.1 years) participated in experiment 1. Twelve participants (11 males and one female, with a mean age of 23.5 years) took part in experiment 2. All participants had normal hearing and normal or corrected-to-normal visual acuity and were naive as to the purpose of the experiment. Participants were paid for taking part in the experiment and informed consent was obtained. This experiment was approved by the ethics committee of the Tokyo Institute of Technology.

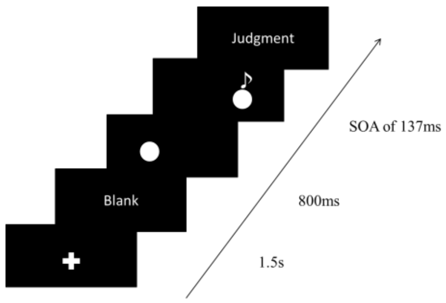
2.2 Apparatus and Stimuli

All TOJ task experiments were conducted in a dark and

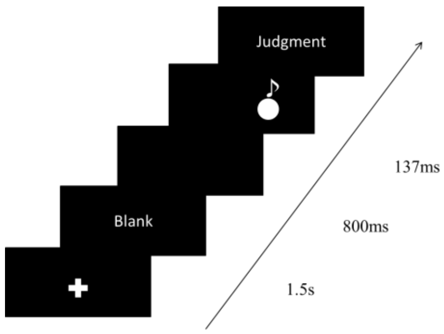
第26回自律分散システム・シンポジウム（2014年1月23日～24日・東京）

soundproof room. Visual stimulation was provided by a 27-inch LCD display (Samsung S27A950D) with a screen resolution of 1920□1080 pixels, and a refresh rate of 120 Hz. The display was operated by a PC workstation (Mac pro, 3.2GHz Quad-Core Intel Xeon, ATI Radeon HD 5770 graphic card, 1GB GDDR5 memory) and placed in front of the subjects. Their head position was fixed by a chin rest at a viewing distance of 100 cm. A white cross of 2 cm in length was displayed as a fixation point in the center of the screen. Visual stimuli consisted of one or two white disks 3.2 cm in diameter on a black background. The visual angle was 2.8° for the single stimulus and 5.6° for the two stimuli. Sound stimuli were presented as mono sounds (65dB, 1,000Hz) delivered via two speakers (MM-SPWD3BK, Sanwa supply). The speakers were located on top of the screen. These visual and auditory stimuli were developed and operated by a computer program (Matlab and Psychtoolbox-3).

(A) Apparent motion condition



(B) Normal condition



(C) Random-order presentation

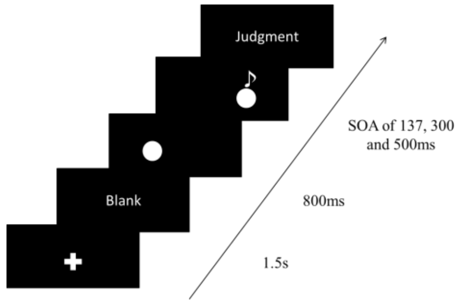


Fig. 1 Schematic illustration of experiment 1 and experiment 2. The two conditions in experiment 1: Apparent motion condition (A) and Normal condition (B). Random-order presentation (C) in experiment 2.

2.3 Procedure

In experiment 1, the participant sat on a chair in front of the stimulation and a constant head position was maintained by chin-rest. The audio-visual TOJ tasks were performed over two sessions with visual stimuli: TOJ task under apparent motion condition and normal condition. Figure 1 illustrates the procedure for experiment 1. In the apparent motion condition (Fig. 1(A)), each trial started with the fixation cross for 1.5 seconds, and a dark blank screen was followed for 800 ms. Next, one white circle for the first visual stimulus showed up for 30 ms and after 137 ms as Stimulus onset asynchrony (SOA), the second stimulus was presented for 30 ms [8]. To assess the temporal discrimination of the auditory and visual stimuli pairs, one brief sound (30ms) as an auditory stimulus was presented with the second visual stimulus. The subjects were instructed to conduct a TOJ task between the second visual frame and the brief sound. The onset time of the auditory stimulus paired with visual stimulus was randomly selected from the following SOA values: -120, -90, -60, -30, 0, +30, +60, +90, and +120 ms (where the negative values indicate that the auditory stimulus preceded the visual stimulus). Then the participant made a forced-choice judgment with respect to the order between the audio-visual stimuli by answering the question ‘which one was first?’ as question mark. The answers consisted of ‘light first’ which was chosen by pressing the Z key and ‘sound first’ which corresponded to the X key. As a way to answer, ‘light first’ was selected when the flash was ahead of the sound, and vice versa with ‘sound first’. In normal condition (Fig. 1(B)), the procedure of single flash condition was the same as TOJ task under apparent motion condition. However, only the second frame in apparent motion condition was shown in this session, so the first visual frame was not presented. Then, the procedure for evaluating the temporal discrimination between sound and flash, and the SOA values were the same as those used for apparent motion condition. The experiment 1 consisted of 270 trials (2 visual conditions × 9 audiovisual SOAs × 15 repeats) with counterbalanced order. Participants performed 27 trials (9 audiovisual SOAs × 3 repeats) as one block for each condition.

I In experiment 2, apparatus, stimuli, and procedure were the same as in experiment 1, with the following exceptions. In experiment 2 only the apparent motion condition was studied. Participants conducted the TOJ tasks with SOAs between the visual stimuli of 137 ms, 300 ms and 500ms presented in a random order. Timing of the auditory stimulus relative to the second flash was the same as in experiment 1. The participants were instructed to judge the order of the second visual frame and the brief sound. The experiment 2 consisted of 432 trials (3 visual conditions × 9 audiovisual SOAs × 16 repeats) with counterbalanced order. Participants performed 54 trials (3 visual conditions × 9 audiovisual SOAs × 2 repeats) as one block for each condition and only the data of apparent motion was calculated in experiment 2. The practice of each experiment was conducted and the total performance took about one and a half hours in each experiment.

Prior to the experimental session, we examined whether the participants perceived motion between two flashes and also confirmed that the motion was perceived during the experiment after the experimental session.

2.4 Data analysis

The ratio of the answers indicating the earlier presentation of the auditory stimulus was calculated for each SOA. We conducted logistic regressions using a generalized linear model with the ratio data of each experiment [1]. The following equation was applied to the regression analysis:

$$y = \frac{1}{1 + e^{\frac{(\alpha - x)}{\beta}}} \quad (1)$$

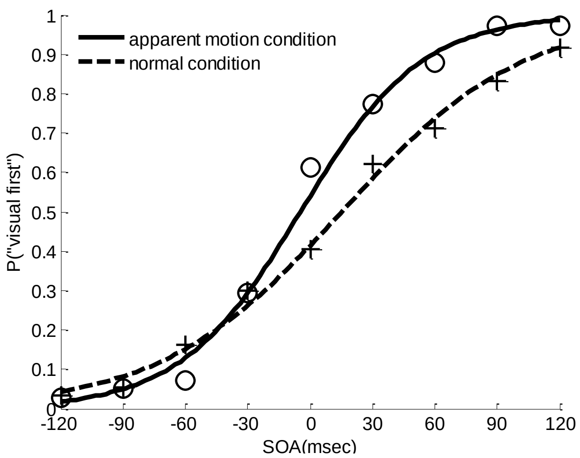
Where α represents the estimated PSS, x denotes SOA, and β is related to JND as shown in the following:

$$JND = \frac{X_{75} - X_{25}}{2} = \beta \log 3 \quad (2)$$

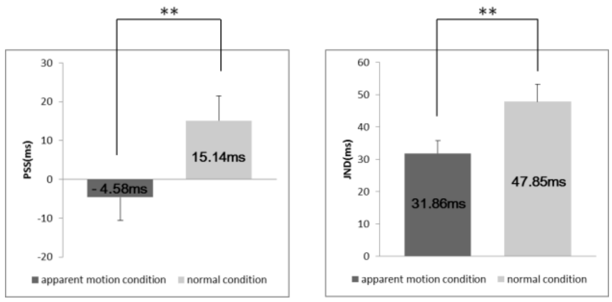
where X_p represents the SOA with p percent of 'auditory first' responses.

We determined the JND and PSS values for each participant using regression analyses (Equation (1) and (2)) and processed the data statistically to obtain mean and standard error values.

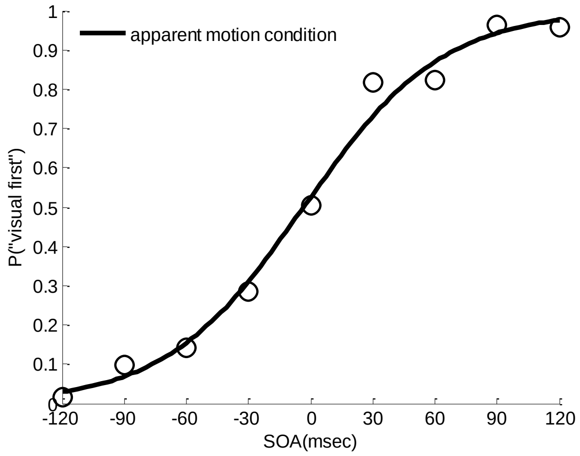
(A)



(B)



(C)



(D)

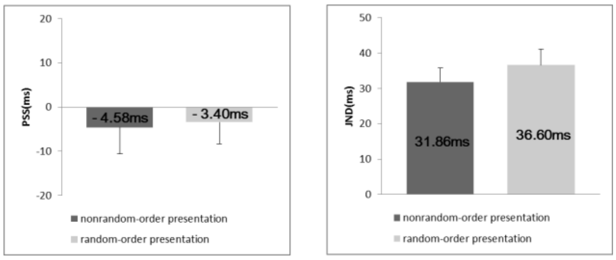


Fig. 2 The results from experiment 1 and experiment 2. (A), (C) Psychometric curves fitted to the distribution of the mean TOJ data in Experiment 1 and 2. (B), (D) Mean PSSs and JNDs in the apparent motion condition and normal condition and random-order presentation condition in Experiment 1 and 2. Error bars represent the standard error of the means.

3. Results

3.1 Experiment 1

The results of two participants were excluded because they were not perceived as continuous motion. Fig. 2 presents the results of experiment 1. As shown in Fig. 2(B), the PSS under the normal condition was a positive value, 15.14 ms (SE = 6.31), but the PSS under apparent motion condition shifted to a negative value, -4.58 ms (SE = 5.99). The PSS of the negative value indicates that the audiovisual stimulus pairs were perceived as simultaneous when the auditory stimuli preceded the visual stimuli. Paired t-test on PSSs indicated significant deference between TOJ task under apparent motion condition and normal condition ($t(11) = -3.46$, $P < 0.01$). Besides, the JND of apparent motion condition was smaller than normal condition (see Fig. 2(B)), and the JND sizes were 31.86 ms (SE = 3.91) and 47.85 ms (SE = 5.45) respectively. A significant difference in paired t-test was observed for the JNDs between them ($t(11) = -3.94$, $P < 0.01$).

3.2 Experiment 2

In experiment 2, Participants performed TOJ task under random order presentation condition and only the results of apparent motion condition were extracted. All participants were perceived as continuous motion and the PSS and JND were computed as in experiment 1. Fig. 2 shows the results of experiment 2, Fig. 2(D) present the results of PSS and JND in experiment 2. The PSS and JND of the apparent motion condition under experiment 2 was almost the same as the results of apparent motion condition in experiment 1. Unpaired t-test on PSSs and JNDs of the TOJ tasks with apparent motion was not significantly different between experiment 1 and experiment 2 ($t(22) = -0.29$, $P = 0.78$, $t(22) = -0.09$, $P = 0.93$).

4. Discussion

We found new evidence that there is a correlation between the visual apparent motion and audiovisual temporal perception. The results of experiment 1 shows that the PSS of normal condition is similar to previous studies, which was usually shifted toward the visual-lead stimuli [1, 9], but the PSS of apparent motion condition was shifted to a negative value, which means that visual processing was faster. In particular, the results of experiment 2 that eliminated the effect of prediction were not different from the result of apparent motion condition in experiment 1. We discuss our new findings to show that the visual apparent motion influences on audiovisual temporal perception.

The apparent motion of the present study has resulted in faster visual processing. Previous studies have reported that PSS on audiovisual simultaneity usually shifts toward a visual-lead stimulus, so simultaneity is maximally perceived if light comes slightly before sound [1, 9]. In addition, many studies have shown that audition dominates vision and sound attracts or captures visual events in the time dimension [5, 6]. However, in this study, with apparent motion stimuli temporal processing of vision was faster. Therefore, our findings suggest that visual motion stimulation contributes to faster visual processing.

With respect to temporal resolution, we found visual apparent motion resulted in the higher temporal discrimination. JND is known as the range of 30-60 ms in audiovisual TOJ task [10, 11]. Previous studies have reported that the different temporal resolution depends on the combination of sensory information and they can change according to a variety of factors such as stimulus intensity and attention [2, 3]. However, the apparent motion shows the higher temporal resolution compared with normal condition. Therefore, we suggest visual apparent motion contributes to higher temporal discrimination.

Also, although participants could not predict the presence or absence of apparent motion the results showed that apparent motion was equivalently processed regardless of prediction. With respect to visual prediction and attention, when subjects know the specific time that targets appear specific attention can be allocated [12]. It is known that the predictable and anticipant information improves the temporal resolution and temporal sensitivity [13]. The attention modulates neural activity and allocates a faster time course for motion processing [14]. But the results of unpredictable apparent motion did not differ from that of predictable apparent motion. Therefore, it is considered to not be the effect of prediction and intention as top-down factors.

Our findings lead to new evidence that motion perception affects temporal perception in audiovisual processing. In unisensory processing, some researchers has been reported that visual motion was perceived faster than non-motion information [15]. However, it remained a need to examine whether motion perception influences temporal perception in multisensory processing or not. What mechanisms contribute to the results which apparent motion affects temporal perception? It may be due to peculiar motion perception mechanisms in primates. Although two discrete stimuli are presented we perceive motion by appropriate spatiotemporal intervals. This phenomenon indicates automatic binding property by motion perception mechanisms in primates. Our findings raise the possibility that the binding property between two flashes influences the binding with a brief sound.

References

[1] Vroomen J, Keetels M (2010) Perception of intersensory synchrony: A tutorial review. Attention, Perception, & Psychophysics 72 (4): 871–884.

[2] Spence C, Shore DI, Klein RM (2001) Multisensory prior entry. Journal of Experimental Psychology 130: 799–832.

[3] Spence C, Parise C (2010) Prior-entry: A review. Consciousness and Cognition 19: 364–379.

[4] Fendrich R, Corballis PM (2001) The temporal cross-capture of audition and vision. Perception & Psychophysics 63(4): 719–725.

[5] Morein-Zamir S, Soto-Faraco S, Kingstone A (2003) Auditory capture of vision: Examining temporal ventriloquism. Cognitive Brain Research 17: 154–163.

[6] Dawson M (1991) The How and Why of What Went Where in Apparent Motion: Modeling Solutions to the Motion Correspondence Problem. Psychological Review 98(4): 569–603.

[7] Ramachandran VS, Anstis SM (1986) The perception of apparent motion. Scientific American 254(6): 102–109.

[8] Harrar V, Winter R, Harris LR (2008) Visuotactile apparent motion. Perception & Psychophysics 70(5): 807–817.

[9] Zampini M, Guest S, Shore DI, Spence C (2005) Audio-visual simultaneity judgments. Perception & Psychophysics 67(3): 531–544.

[10] Morein-Zamir S, Soto-Faraco S, Kingstone A (2003) Auditory capture of vision: Examining temporal ventriloquism. Cognitive Brain Research 17: 154–163.

[11] Keetels M, Vroomen J (2005) The role of spatial disparity and hemifields in audio–visual temporal order judgments. Experimental Brain Research 167: 635–640.

[12] Coull JT, Nobre AC (1998) Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. Journal of Neuroscience, 18(18): 7426–7435.

[13] Petrini K, Russell M, Pollick F (2009) When knowing can replace seeing in audiovisual integration of actions. Cognition 110: 432–439.

[14] Busse L, Katzner S, Treue S (2008) Temporal dynamics of neuronal modulation during exogenous and endogenous shifts of visual attention in macaque area MT. Proceedings of the National Academy of Sciences of the United States of America, 105(42): 16380–16385.

[15] Nijhawan R (1994) Motion extrapolation in catching. Nature 370: 256–257.