Temporal Order Perception of Sensory Stimuli under Motion Perception in Audiovisual Processing

Jinhwan Kwon[1†] , Ken-ichiro Ogawa[1] and Yoshihiro Miyake[1]

[1]Department of Computational Intelligence and Systems Science,
Tokyo Institute of Technology, Tokyo, Japan
(Tel : +81-45-924-5656; E-mail: kwon@myk.dis.titech.ac.jp,
ogawa@dis.titech.ac.jp, miyake@dis.titech.ac.jp)

Abstract: Motion perception has an important role in a dynamic environment. However, the mechanism of perception of external stimuli under motion perception is not fully clarified. This study investigated the relationship between temporal order perception and motion perception on audiovisual processing. Participants performed audiovisual temporal order judgment (TOJ) task under apparent motion condition and non-apparent motion condition in Experiment 1 and also under random-order presentation condition in Experiment 2. Our result shows that the perceived order and temporal resolution between audiovisual stimuli is different between apparent motion perception and non-apparent motion perception. Besides, the perceived order during apparent motion perception was processed regardless of prediction. In particular, motion perception and exogenous attention are closely related. The relationship shows that motion perception may be exogenously processed by bottom-up signals of external stimulus.

Keywords: Motion Perception, Apparent Motion, Temporal Order, Audiovisual Simultaneity

1. INTRODUCTION

We perceive and interpret external sensory stimuli to establish flexible interaction with the environment. In particular, motion perception has an important role in a dynamic environment. However, the mechanism of temporal perception under motion perception is not fully elucidated. This study investigated the temporal order perception of external sensory stimuli under motion perception on audiovisual integration.

Motion information is a very important factor in the dynamic environment. We interact with the dynamic environment by perceiving and interpreting external information, especially motion information. Interestingly, it is reported that there exists a temporal difference between the presentation of external sensory stimuli and the perception of the stimuli [1, 2]. External sensory stimuli are perceived with different delays due to transmission time through air and neural transmission time [3]. However, we perceive motion information smoothly in real time in a dynamic environment.

With respect to motion perception, apparent motion is fundamental unit. It is reported that temporal and spatial characteristics for visual motion perception are determined on human perceptual system [4-8]. Especially, even despite two discrete stimuli, motion is perceived by the appropriate spatiotemporal interval [4-8]. This shows some perceptual frame for perceiving continuous motion. Many researchers have reported that two visual stimuli are perceived as a continuous motion when the stimulus onset asynchrony (SOA) of the visual stimuli is within a range of 50 to 150 ms [6, 9, 10]. Conversely, the visual stimuli are perceived as successive beyond an ISOI of 300 ms. This characteristic allows us to perceive the discrete stimuli to be perceived as continuous and it is applied to movies and television (e.g., 24 fps in films, 30 fps in television).

When any event occurs, multisensory information on the event is transmitted through air and is perceived by each sensory receptor. Then, we integrate and interpret the multiple sensory information as an external event. However, because the transmission time through air and neural transmission time are different in kind of sensory stimuli, it is reported that the perceived order of the sensory stimuli is different in the brain. Temporal order judgment (TOJ) task is known as a psychophysical study to examine the temporal order perception of external stimuli in multisensory processing [3]. However, although the temporal order perception of external stimuli is quite essential in a dynamic environment, the reports involved in this situation are insufficient [1-3]. We therefore focused on the temporal order perception of during motion perception.

The purpose of the present study is to investigate the temporal order perception of external sensory stimuli on motion perception. We examined two types of TOJ task experiments. In Experiment 1, we examined whether visual apparent motion affect audiovisual TOJ task. Participants performed audiovisual TOJ task between the apparent motion condition and the single flash condition. In Experiment 2, we examined whether motion perception is affected by timing prediction. We therefore confirmed the effect of the timing prediction on motion perception by presenting the stimuli perceived as motion (SOA of 137 ms) and non-motion (SOA of 300 and 500 ms) in a random order.
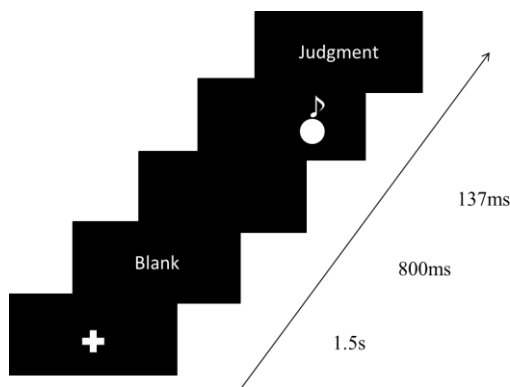
2. METHOD

2.1 Participants

Sixteen participants (15 males and one female, with a mean age of 24.3 years) participated in experiment 1. Twelve participants (10 males and two females, with a mean age of 23.9 years) took part in experiment 2. All participants had normal hearing and normal or

(A) Apparent motion condition
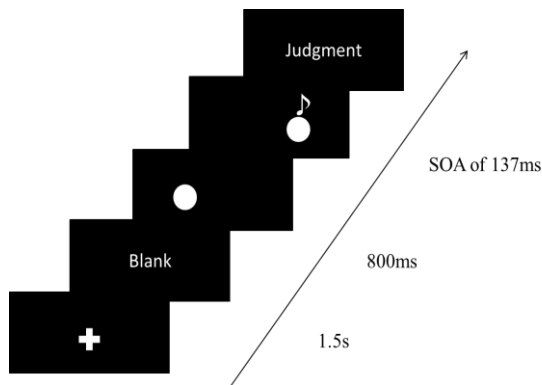
(B) Single flash condition

(C) Apparent motion condition and Non-apparent motion condition in random-order presentation
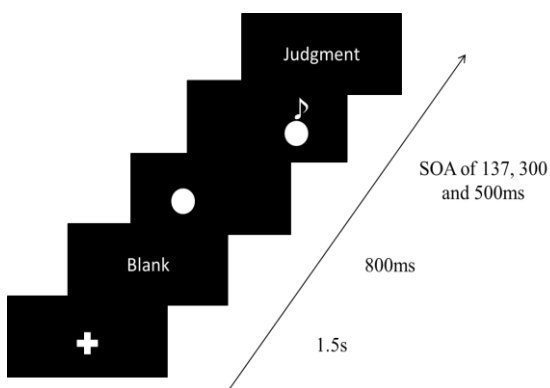

Fig. 1 Schematic illustration of experiment 1 and experiment 2. The two conditions in experiment 1: Apparent motion condition with SOA of 137 ms (A) and Normal condition with single flash (B). Three conditions in in experiment 2: Apparent motion condition with SOA of 137 ms and Non-apparent motion condition with SOA of 300 and 500 ms.

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 written 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 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).

2.3 Procedure

In experiment 1, the participant sat on a chair facing the stimulus, and a constant head position was maintained by the chin rest. The audiovisual TOJ tasks were performed over two sessions with visual stimuli: in the apparent motion condition and in the single flash condition. Figure 1 illustrates the procedure for experiment 1. In the apparent motion condition (Fig. 1(A)), each trial began with display of the fixation cross for 1.5 seconds, followed by a dark blank screen for 800 ms. Next, one white circle for the first visual stimulus was displayed for 30 ms; then with a SOA of 137 ms, the second stimulus was presented for 30 ms [11]. To assess the temporal discrimination of the auditory and visual stimulus pairs, one brief sound (30ms) as an auditory stimulus was presented at different times relative to 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 a 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 of the audiovisual stimuli by answering the question 'which one was first?' The answers consisted of 'light first,' which was chosen by pressing the Z key, and 'sound first,' which corresponded to the X key. The response 'light first' was selected when the flash was ahead of the sound, and vice versa for 'sound first.' In the single flash condition (Fig. 1(B)),
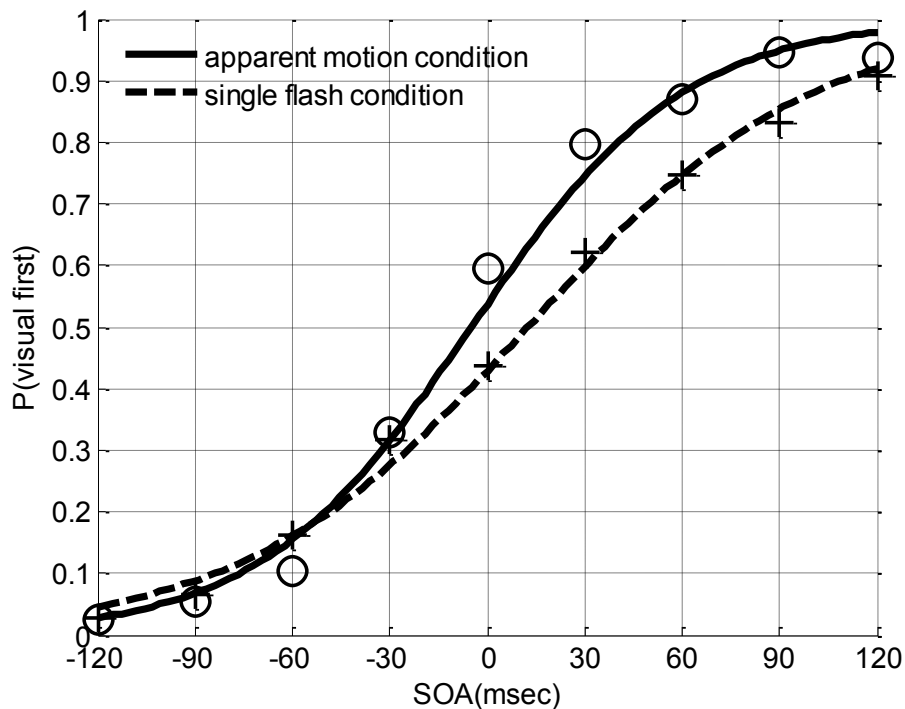
Fig. 2 The results of apparent motion condition and single flash condition in Experiment 1. Psychometric curves fitted to the distribution of the mean TOJ data in Experiment 1. The 50% crossover point is taken as PSS, which is maximally simultaneous about the temporal order in SOA. JND is defined as a half between the 25% and 75% point in SOA, which is used as an indicator that discriminates the temporal resolution in cross-modality.

the procedure for the single flash condition was the same as that for the TOJ task in the apparent motion condition. However, only the second frame in the apparent motion condition was shown in this session; the first visual frame was not presented. Then, the same procedure for evaluating the temporal discrimination between sound and flash, and same SOA values were used as in the apparent motion condition. Experiment 1 consisted of 270 trials (2 visual conditions × 9 audiovisual SOAs × 15 repeats) in counterbalanced order. Participants performed 27 trials (9 audiovisual SOAs × 3 repeats) as one block for each condition. 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

Point of subjective simultaneity (PSS) and just noticeable difference (JND) are used as the methods of measurement. The PSS represents the interval between the application of stimuli to two senses at which both are perceived by the senses to occur the same time, which makes it possible to detect which sensory information was captured early or late. The JND has been used as an indicator that discriminates the temporal resolution in cross-modality. 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 [12]. The following equation was applied to the regression analysis:
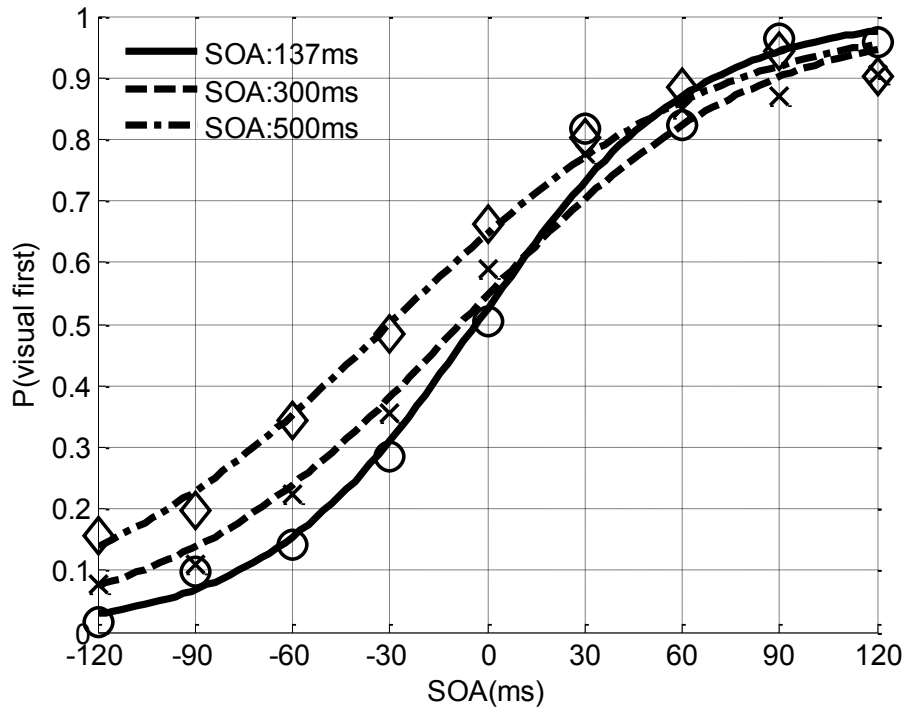
Fig. 3 The results of apparent motion condition (SOA : 137 ms) and non-apparent motion condition (SOA : 300 and 500 ms) in Experiment 2. Psychometric curves fitted to the distribution of the mean TOJ data in Experiment 2. The 50% crossover point is taken as PSS, which is maximally simultaneous about the temporal order in SOA. JND is defined as a half between the 25% and 75% point in SOA, which is used as an indicator that discriminates the temporal resolution in cross-modality.

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

where α represents the estimated PSS, x denotes SOA. y indicates the responses of visual first or auditory first. JND is defined as shown in the following:

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

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 (Eqs. (1) and (2)) and processed the data statistically to obtain mean and standard error values.

3. RESULTS

3.1 Experiment 1
Fig. 2 presents the results of Experiment 1. As shown in Table 1, the PSS in the single flash condition had a positive value, 12.47 ms (SE = 6.45), but the PSS in the apparent motion condition shifted to a negative value, –4.90 ms (SE = 5.84). The PSS of the negative value indicates that the audiovisual stimulus pairs were perceived as simultaneous when the auditory stimuli preceded the visual stimuli. A paired t-test of PSSs indicated significant difference between the TOJ task in the apparent motion condition and that in the single flash condition (t(15) = –2.33, P < 0.05). In addition, the JND in the apparent motion condition was smaller than that in the single flash condition (see Table 1), and the JND values were 35.72 ms (SE = 3.96) and 48.23 ms (SE = 5.17), respectively. A significant difference between the JNDs was observed in the paired t-test (t(15) = –3.57, P < 0.01).

Table 1 The results of Experiment 1.

	Apparent motion condition	Single flash condition	t-test
PSS	- 4.90	12.47	*
JND	35.72	48.23	**

* : P < .05, ** : P < .01, n.s. : not significant

3.2 Experiment 2

In Experiment 2, the participants performed the TOJ task with the three intervals between the visual stimuli in a random order. The PSSs and JNDs were computed as in Experiment 1. Fig. 3 shows the results of Experiment 2, Table 2 show the results for PSSs and indicate JNDs in Experiment 2. The PSS shifted toward a sound-lead stimulus in the three condition. Especially, PSS shifted toward a sound-lead stimulus as the temporal interval between two flashes is longer (Table 2). Moreover, the JND in the apparent motion perception was smaller than that in the other intervals (Table 2). A repeated-measures analysis of variance (ANOVA) of the PSSs showed the significant main effect of the temporal interval, $F_{(2, 23)} = 13.07$, $p < 0.002$. Besides, a repeated-measures ANOVA of the JNDs revealed a significant main effect of the temporal interval, $F_{(2, 23)} = 4.69$, $p < 0.041$. In particular, the values of PSS and JND in the apparent motion condition in Experiment 2 were almost the same as those of the apparent motion condition in Experiment 1 (Table 3). An unpaired t-test of PSSs and JNDs of the TOJ tasks in the apparent motion condition indicated no significant difference between Experiment 1 and Experiment 2 ($t(26) = –0.11$, $P = 0.92$, $t(26) = –0.12$, $P = 0.91$).

Table 2 The results of Experiment 2.

	137 ms	300 ms	500 ms	F-test
PSS	-3.40	-8.42	-30.07	**
JND	36.60	48.92	54.30	*

* : P < .05, ** : P < .01, n.s. : not significant

Table 3 Comparison of results of apparent motion condition between Experiment 1 and Experiment 2.

	Apparent motion condition		t-test
	Experiment 1	Experiment 2	
PSS	- 4.90	-3.40	n.s.
JND	35.72	36.60	n.s.

* : P < .05, ** : P < .01, n.s. : not significant

4. DISCUSSION

The result of Experiment 1 shows that the PSS in the single flash condition was shifted toward the visual-lead stimulus, but the PSS in the apparent motion condition was shifted to a sound-lead stimulus. Besides, the JND in the apparent motion condition was smaller than that in the single flash condition. In Experiment 2, we investigated the effect of timing prediction by modulating the temporal interval between two flashes in a random order. The result of Experiment 2 shows that there was a significant difference from one temporal interval to others in the PSS and JND. However, the result of apparent motion condition in Experiment 2 was not different from that in the apparent motion condition in Experiment 1. These results indicate that visual

apparent motion influences audiovisual temporal order perception.

In detail, the result of Experiment 1 shows that the PSS in the single flash condition is similar to that of previous studies. The previous studies reported that the PSS was usually shifted toward the visual-lead stimulus [14-17]. Therefore, auditory stimulus is slightly perceived before visual stimulus. However, the PSS in the apparent motion condition was shifted to a sound-lead stimulus. This indicates that visual stimulus is more slightly perceived before auditory stimulus and the perceived order of audiovisual stimuli is different from the result in the apparent motion condition and the single flash condition. The JND in the apparent motion condition was smaller than that in the single flash condition. This shows that apparent motion perception results in higher temporal resolution and is sensitive to temporal asynchrony of external stimuli [3].

Next, the result of Experiment 2 shows that PSS and JND were different from one to the others. In particular, the PSS in the SOA of 300 and 500 ms turned into a sound-lead stimulus more than that in the SOA of 137 ms between two flashes. This result indicates that the SOA of 300 and 500 ms leads to stronger visual attention than the SOA of 137 ms because the unpredictable temporal interval becomes longer in the SOA of 300 and 500 ms. Moreover, the JND was smaller in the apparent motion perception. This shows that temporal resolution was higher in apparent motion perception compared with non-apparent motion perception.

However, although it is known that the predictable and anticipant information improves the temporal resolution and temporal sensitivity [18], there is no difference between the results in predictable apparent motion (Experiment 1) and in unpredictable apparent motion (Experiment 2). This suggests that apparent motion perception is not affected by timing prediction. With respect to prediction and intention, there is a possibility that exogenous attention known as transient attention activated in apparent motion perception. The exogenous attention shows an automatic and involuntary system by external stimulation that rises and decays quickly and activates from maximally about 100-120 ms and is effective up to 300 ms. [19-24]. In particular, the temporal intervals for apparent motion perception are similar to the maximal intervals for the activation of exogenous attention. Therefore, motion perception seems to be closely connected to exogenous attention. This suggests that motion perception may be exogenously processed by bottom-up signals involved in external stimuli.

We investigated the temporal order perception of external stimuli under motion perception. The results of Experiment 1 and 2 show that the temporal order perception and the temporal resolution of audiovisual stimuli during apparent motion perception are different from these during non-apparent motion perception, and the temporal order perception during apparent motion perception was equivalently processed regardless of

temporal prediction. It is also suggested that motion perception and exogenous attention are closely related, and therefore motion perception may be exogenously processed by bottom-up signals from external stimuli.

ACKNOWLEDGMENT

REFERENCES

[1] C. Spence, D.I. Shore and R.M. Klein, "Multisensory Prior Entry," *Journal of Experimental Psychology* Vol. 130, No. 4, pp. 799–832, 2001.

[2] C. Spence and C. Parise, "Prior-entry: A Review," Consciousness and Cognition, Vol. 19, No. 1, pp.364–379, 2010.

[3] J. Vroomen and M. Keetels, "Perception of Intersensory Synchrony: A Tutorial Review," *Attention, Perception, & Psychophysics*, Vol. 72, No. 4, pp.871–884, 2010.

[4] A. Larsen, J.E. Farrell and C Bundesen, "Short- and Long-Range Processes in Visual Apparent Movement," *Psychological Research*, Vol. 45, No. 1, pp.11–18, 1983.

[5] O.J. Braddick, K. H. Ruddock, M.J. Morgan and D. Marr, "Low-Level and High-Level Processes in Apparent Motion," *Philosophical transactions Royal Society London B*, Vol. 290, No. 1038, pp.137-151, 1980.

[6] M.R.W. Dawson, "The How and Why of What Went Where in Apparent Motion: Modeling Solutions to The Motion Correspondence Problem," *Psychological Review*, Vol. 98, No. 4, pp.569-603, 1991.

[7] S. Grossberg and M.E. Rudd, "Cortical Dynamics of Visual Motion Perception: Short-range and long-range Apparent Motion," *Psychological review*, Vol. 99, No. 1, pp.78-121, 1992.

[8] A.B. Watson, A.J. Ahumada and J.E. Farrell, "Window of Visibility: A Psychophysical Theory of Fidelity in Time-sampled Visual Motion Displays," *Optical Society of America*, Vol. 3, No. 3, pp.300-307, 1986.

[9] S. Getzmann, "The effect of Brief Auditory Stimuli on Visual Apparent Motion," *Perception*, Vol. 36, pp.1089–1103, 2007.

[10] T.Z. Strybel, C.L. Manligas, O. Chan and D.R. Perrott "A Comparison of The Effects of Spatial Separation on Apparent Motion in The Auditory and Visual Modalities," *Perception & Psychophysics*, Vol. 47, No. 5, pp.439–448, 1990.

[11] V. Harrar, R. Winter and L.R. Harris, "Visuotactile Apparent Motion," *Perception & Psychophysics*, Vol. 70 No.5, pp.807–817, 2008.

[12] D.J. Finney, "Probit analysis: A Statistical Treatment of The Sigmoid Response Curve," *Cambridge Univ. Press*, 1952.

[13] P. Jakowski, F. Jaroszyk and D. Hojan-Jezierska, "Temporal-order Judgments and Reaction Time for Stimuli of Different Modalities," *Psychological Research*, Vol. 52, No. 1, pp.35–38, 1990.

[14] J. Lewald, R. Guski, "Cross-modal Perceptual Integration of Spatially and Temporally Disparate Auditory and Visual Stimuli," *Cognitive Brain Research*, Vol. 16, No. 3, pp.468–478, 2003.

[15] M. Zampini, S. Guest, D.I. Shore and C. Spence, "Audio-visual Simultaneity Judgments," *Perception & Psychophysics*, Vol. 67, No.3, pp.531–544, 2005.

[16] M. Kanabus, E. Szelg, E. Rojek and E. Pöppel, "Temporal Order Judgement for Auditory and Visual Stimuli," *Acta Neurobiologiae Experimentalis*, Vol. 62, No. 4, pp.263–270, 2002.

[17] K. Petrini, M. Russell and F. Pollick, "When Knowing Can Replace Seeing in Audiovisual Integration of Actions," *Cognition*, 110, pp.432–439, 2009.

[18] H.J. Müller and P.M. Rabbitt, "Reflexive and Voluntary Orienting of Visual Attention: Time Course of Activation and Resistance to Interruption," Journal of Experimental Psychology: *Human Perception and Performance*, Vol. 15, No. 2, pp. 315–330, 1989.

[19] E. Hein, B. Rolke, R. Ulrich, "Visual Attention and Temporal Discrimination: Differential Effects of Automatic and Voluntary Cueing," *Visual Cognition*, Vol. 13. No. 1, pp.29-50, 2006.

[20] S. Ling, M. Carrasco, "Sustained and Transient Covert Attention Enhance The Signal via Different Contrast Response Functions," *Vision Research*, Vol. 46(8-9), pp.1210-1220, 2006.

[21] T. Liu, S.T. Stevens, M. Carrasco, "Comparing The Time Course and Efficacy of Spatial and Feature-based Attention," *Vision Research*, Vol. 47, No. 1, pp.108-113, 2007.

[22] K. Nakayama, M. Mackeben, "Sustained and Transient Components of Focal Visual Attention," *Vision Research*, Vol. 29, No. 11, pp.1631-1647, 1989.

[23] R.W. Remington, J.C. Johnston, S. Yantis, "Involuntary Attentional Capture by Abrupt Onsets," *Perception & Psychophysics*, Vol. 51, No. 3, pp.279-290, 1992.