2014 ICME International Conference on Complex Medical Engineering
June 26-29, 2014, Taipei, Taiwan

Temporal Processing for Visual Motion Information in Audiovisual Integration

Jinhwan Kwon, Ken-ichiro Ogawa
Department of Computational Intelligence and Systems Science
Tokyo Institute of Technology
4259 Nagatsuda, Midori, Yokohama 226-8502, Japan
email: kwon@myk.dis.titech.ac.jp, ogawa@dis.titech.ac.jp

Taiki Ogata
Research into Artifacts, Center for Engineering(RACE)
The University of Tokyo
5-1-5, Kashiwanoha, Kashiwa-shi, Chiba, 277-8568, Japan
email: ogata@race.u-tokyo.ac.jp

JongHwan Kim
Graduate School of Information Security
Korea Advanced Institute of Science and Technology
291 Daehak-ro(373-1 Guseong-dong), Yuseong-gu, Daejwon 305-701, Korea
email: zzoru@kaist.ac.kr

Yoshihiro Miyake
Department of Computational Intelligence and Systems Science
Tokyo Institute of Technology
4259 Nagatsuda, Midori, Yokohama 226-8502, Japan
email: miyake@dis.titech.ac.jp

***Abstract* - Given a dynamic environment, observing motion is essential for the timing of behavior. Apparent motion clearly represents spatiotemporal characteristics of movement with respect to observing motion. We investigated an effect of visual motion perception on temporal processing in audiovisual integration. Participants performed audiovisual temporal order judgment (TOJ) task under apparent motion condition and normal condition and also random-order presentation condition. As a result, the PSS and JND in the apparent motion condition differed from the results of those in the normal condition regardless of prediction. Therefore, this findings show that motion perception has a different multisensory mechanism relative to non-motion perception. Especially, we suggest that the area MT may contribute to the audiovisual temporal processing with visual motion information.**

Index Terms *– Apparent Motion, Motion Perception, Temporal Perception, Audiovisual Simultaneity*

I. Introduction

The relationship between motion and temporal perception is essential for a dynamic environment. Motion can be separated into temporal and spatial elements in which we observe movement. Observing motion is essential for determining the timing of behavior. But the correlation between motion and temporal perception is not fully clarified concerning multisensory integration. In this study, we focused on the temporal perception during the observation of motion in multisensory integration.

Apparent motion is an important phenomenon in which spatiotemporal characteristics of movement are clearly represented with respect to observing motion. Apparent motion is an optical phenomenon that makes motion appear by an appropriate spatiotemporal interval even despite two discrete stimuli [1], [2], [3], [4] and has been studied in a wide range of fields from psychophysics to brain function [5], [6]. Psychophysical aspects have been studied to elucidate the cognitive or perceptual characteristics of apparent motion and spatiotemporal factors are considered to be important. In the physiological aspect, the response to apparent motion is typically equivalent to the response to real motion in the physiological mechanisms and passes through the dorsal stream associated with motion processing [7], [8].

With respect to temporal perception, the important question is how multiple senses are integrated in the time dimension. In particular, it is important to examine the temporal characteristics between external stimuli and internal perception. Temporal order judgment (TOJ) task is known as a psychophysical study to examine the temporal relationship between external stimuli and internal perception in multisensory processing [9], [10], [11]. 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 [12]. However, although the temporal perception of motion is essential for the dynamic environment, the reports involved in the motion perception are insufficient [12], [13], [14]. Therefore, we focused on the relationship between motion perception and temporal perception through apparent motion and TOJ task.

The purpose of the present study is to investigate an effect of visual motion perception on temporal processing in audiovisual integration. We examined two types of TOJ task experiments. In experiment 1, participants conducted TOJ task in apparent motion condition and normal condition with single flash on audiovisual simultaneity to examine how visual apparent motion affects audiovisual TOJ task. Also, because there remained the specific prediction by constant temporal intervals in apparent motion condition, in experiment 2, we examined supplemental experiment to eliminate the influence of prediction by presenting two visual stimuli in a random order.

II. METHODS

A. Participants

Sixteen participants (15 males and one female, with a mean age of 24.3 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 written informed consent was obtained. This experiment was approved by the ethics committee of the Tokyo Institute of Technology.

B. 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).

C. 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 [15]. 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.

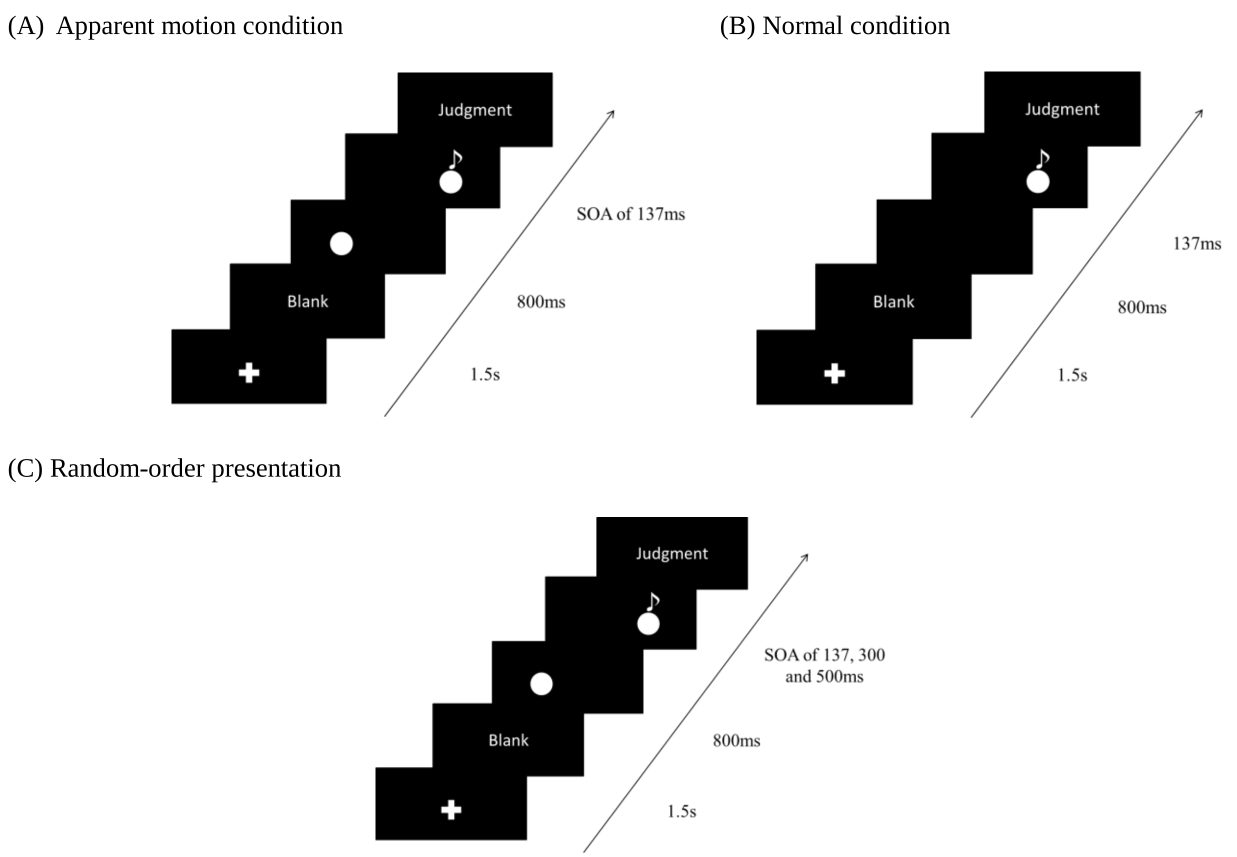

Fig. 1 Schematic illustration of experiment 1 and experiment 2. The two conditions in experiment 1: Apparent motion condition (A) and Normal condition (B). Random-order presentation (C) in experiment 2.

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.

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.

D. 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 [16]. The following equation was applied to the regression analysis:

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

where α represents the estimated PSS, x denotes SOA. 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 (Equation (1) and (2)) and processed the data statistically to obtain mean and standard error values.

III. Results

A. Experiment 1

Fig. 2 presents the results of experiment 1. As shown in Fig. 2(B), the PSS in the normal 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 normal condition ($t(15) = -2.33$, $P < 0.05$). In addition, the JND in the apparent motion condition was smaller than that in the normal condition (see Fig. 2(C)), 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 ($t(15) = -3.57$, $P < 0.01$).

B. Experiment 2

In experiment 2, all participants performed the TOJ task with the intervals between the visual stimuli in a random order, and only the results of the apparent motion condition were extracted. The participants perceived continuous motion, and the PSSs and JNDs were computed as in experiment 1. Fig. 3 shows the results of experiment 2, Fig. 3(B) and 3(C) show the results for PSSs and JNDs in experiment 2. 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. 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 ($t(26) = -0.11$, $P = 0.92$, $t(26) = -0.12$, $P = 0.91$).

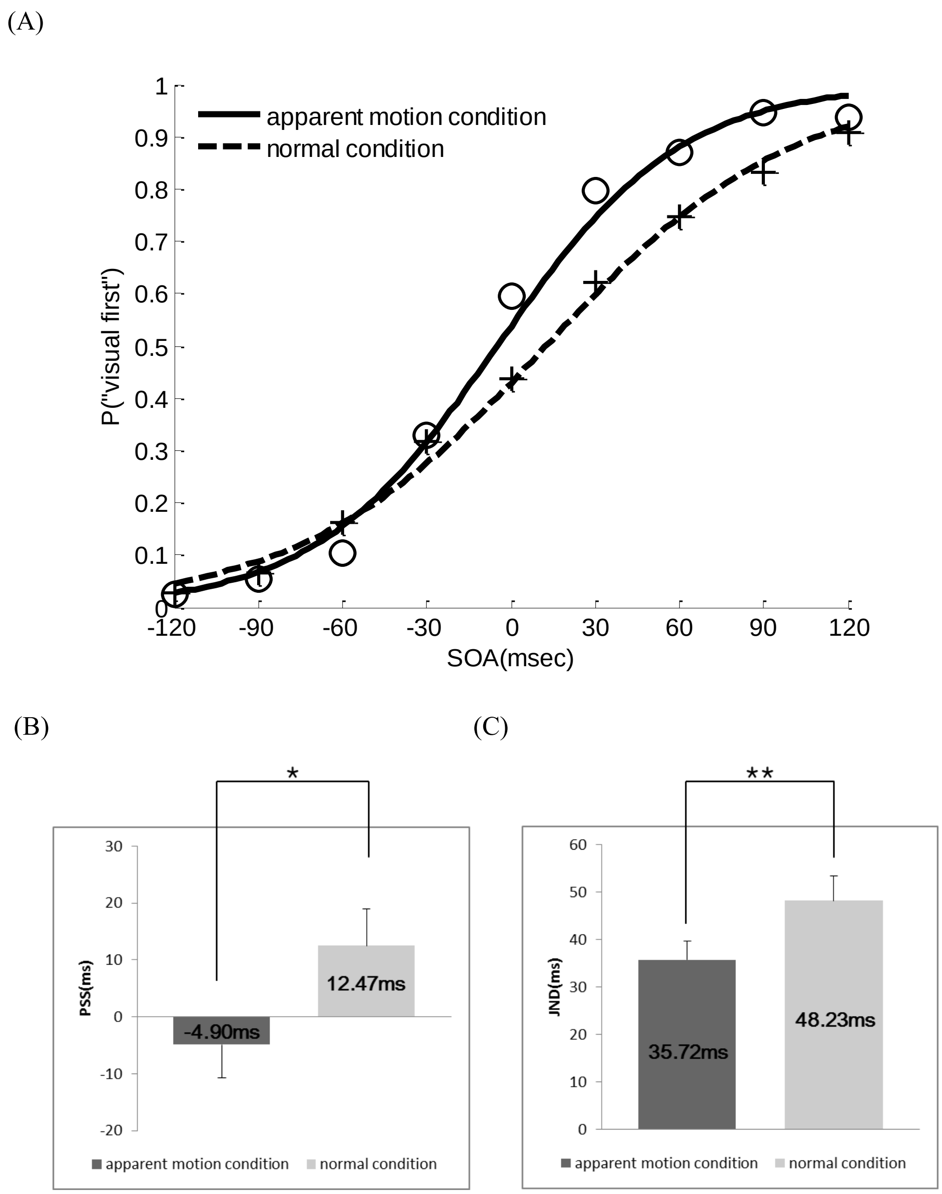

Fig. 2 The results of experiment 1. (A) Psychometric curves fitted to the distribution of the mean TOJ data in Experiment 1. (B), (C) Mean PSSs and JNDs in the apparent motion condition and normal condition in experiment 1. Error bars represent the standard error of the means.

IV. Discussion

We found that the PSS was shifted to a sound-lead stimulus and JND was smaller in apparent motion condition. The sound-lead value of PSS means that visual processing was faster and smaller JND indicates that the temporal resolution between audiovisual stimuli was greater. Also, similar results of PSS and JND were obtained in Experiment 2 regardless of prediction. Previous studies have reported that PSS usually shifts toward visual-lead stimulus within the range of 20-40ms, and therefore simultaneity is maximally perceived if light comes slightly before sound [17], [18], [19], [20]. However, the PSS in apparent motion condition was shifted to a sound-lead stimulus that is the opposite result in normal condition. It is the opposite tendency with audiovisual perceptual characteristics in the temporal domain. With respect to temporal resolution, JND is known as the range of 30-60 ms in audiovisual TOJ task [21], [22], [23]. We however found visual apparent motion resulted in the higher temporal discrimination. This finding suggests that visual motion information contributes to faster visual processing and greater temporal resolution regardless of prediction.

(A)

(B) (C)

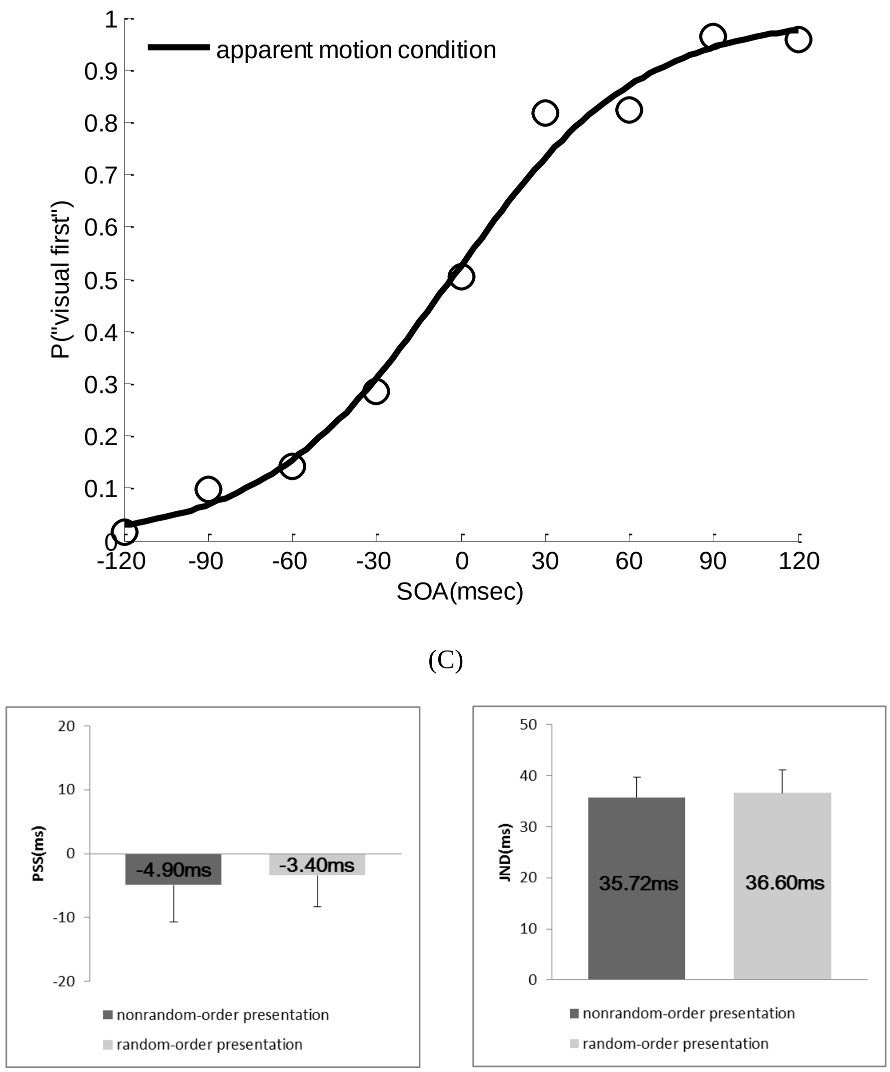

Fig. 3 The results of experiment 2. (A) Psychometric curves fitted to the distribution of the mean TOJ data in Experiment 2. (B), (C) Mean PSSs and JNDs from nonrandom-order presentation (apparent motion condition in experiment 1) and random-order presentation (apparent motion condition in experiment 2). Error bars represent the standard error of the means.

Our findings show that motion perception influences the temporal perception on audiovisual processing. There is a need to examine whether motion information affects the temporal perception in multisensory integration. In unisensory processing, it has been reported that motion was perceived faster than non-motion information [24], [25]. For example, there is a robust illusory phenomenon called flash-lag effect, which motion perceived faster than single flash [24], [26]. When a moving target and a flash are aligned and when they are appeared in the same location the moving target was perceived as more slightly shifted in the direction of motion relative to the flash. The flash-lag effect support our findings but our findings lead to new evidence that motion affects temporal perception in multisensory integration.

With respect to processing of audiovisual integration, functional imaging studies have found the brain areas of audiovisual integration and, especially superior temporal sulcus (STS) have been suggested as an audiovisual association area [27], [28], [29], [30]. However, in recent years, with the growing interest in the multisensory property of motion, some researchers have raised the possibility that there exist different mechanisms by motion information in multisensory processing [31], [32]. So, several functional imaging have also showed the evidences that the audiovisual integration engaged area MT is related to visual motion processing [33], [34].

The area MT in the dorsal stream plays an important role in motion processing [35] and has a function which comprehensively processes the whole motion information of objects by integrating the information of local motion [36]. In addition, the area MT is greatly implicated in precise spatiotemporal encoding with respect to the timing and interval of visual stimuli [37], [38], [39], [40]. Although the area MT has been known as a level in hierarchy of visual processing, since it has recently reported that it is not only the hierarchy of visual pathways and it could be affected by auditory stimulation [34], [41], [42], [43], the reports suggest that the area MT is implicated in multisensory processing. Also, the area MT has the fastest response latencies in visual response in macaque cerebral cortex and TPO corresponding to the STS, which has slower response latencies relative to the area MT [44]. Along with the results of PSS in the present study, we suggest that there exist the difference between the multisensory characteristics of visual motion information and non-visual motion information, and especially the area MT may contribute to the audiovisual temporal processing with motion information.

REFERENCES

[1] V.S. Ramachandran and S.M. Anstis, “The perception of apparent motion,” Scientific American 254(6), pp.102–109, 1986.

[2] S.M. Anstis and D.M. Mackay, “The Perception of Apparent Movement,” Phil. Trans. R. Soc. Lond. B 290, pp.153–168, 1980.

[3] A. Larsen, J.E. Farrell and C Bundesen, “Short- and Long-Range Processes in Visual Apparent Movement,” Psychol Res 45, pp.11–18, 1983.

[4] 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 290, pp.137-151, 1980.

[5] 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.

[6] S. Grossberg, 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.

[7] W.T. Newsome, A. Mikami and R.H. Wurtz, “Motion selectivity in macaque visual cortex. III. Psychophysics and physiology of apparent motion,” Journal of Neurophysiology, 55, pp.1340-1351, 1986.

[8] M.L. Manning, D.C. Finlay and B. Fenelon, “Visual evoked potentials to stimuli in apparent motion,” Vision Research, 28, pp.965-974, 1988.

[9] D.M. Eagleman, “Human time perception and its illusions,” Curr Opin Neurobiol 18, pp.131–136, 2008.

[10] R.B. Ivry and J.E. Schlerf, “Dedicated and intrinsic models of time perception,” Trends Cogn Sci 12, pp.273–80, 2008.

[11] S. Grondin, “Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions,” Atten Percept Psychophys 72, pp.561–582, 2010.

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

[13] C. Spence, D.I. Shore and R.M. Klein, “Multisensory prior entry,” Journal of Experimental Psychology 130, pp.799–832, 2001.

[14] C. Spence and C. Parise, “Prior-entry: A review,” Consciousness and Cognition 19, pp.364–379, 2010.

[15] V. Harrar, R. Winter and L.R. Harris, “Visuotactile apparent motion,” Perception & Psychophysics 70(5), pp.807–817, 2008.

[16] D.J. Finney, “Probit analysis: A statistical treatment of the sigmoid response curve,” Cambridge Univ. Press, 1952.

[17] P. Jakowski, F. Jaroszyk and D. Hojan-Jezierska, “Temporal-order judgments and reaction time for stimuli of different modalities,” Psychol Res 52, pp.35–38, 1990.

[18] M. Zampini, S. Guest, D.I. Shore and C. Spence, “Audio-visual simultaneity judgments,” Perception & Psychophysics 67(3), pp.531–544, 2005.

[19] M. Kanabus, E. Szelg, E. Rojek and E. Pöppel, “Temporal order judgement for auditory and visual stimuli,” Acta Neurobiol. Exp 62, pp.263–270, 2002.

[20] M. Zampini, S. Guest, D.I. Shore and C. Spence, “Audio–visual simultaneity judgments,” Perception & Psychophysics 67, pp.531–544, 2005.

[21] S. Morein-Zamir, S. Soto-Faraco and A. Kingstone, “Auditory capture of vision: Examining temporal ventriloquism,” Cognitive Brain Research 17, pp.154–163, 2003.

[22] M. Keetels and J. Vroomen, “The role of spatial disparity and hemifields in audio–visual temporal order judgments,” Experimental Brain Research 167, pp.635–640, 2005.

[23] M. Zampini, D.I. Shore and C. Spence, “Audiovisual temporal order judgments,” Experimental Brain Research 152, pp.198–210, 2003.

[24] R. Nijhawan, “Motion extrapolation in catching,” Nature 370, pp.256–257, 1994.

[25] O, Hikosaka, S. Miyauchi and S. Shimojo, “Focal visual attention produces illusory temporal order and motion sensation,” Vision Research 33(9), pp.1219–1240, 1993.

[26] J. Vroomen and B.D. Gelder, “Temporal ventriloquism: Sound modulates the flash-lag effect,” Journal of Experimental Psychology: Human Perception & Performance, 30, pp.513-518, 2004.

[27] B.E. Stein and T.R. Stanford, “Multisensory integration: current issues from the perspective of the single neuron,” Nature Reviews Neuroscience vol. 9, pp.255–266, 2008.

[28] A.A. Ghazanfar and C.E. Schroeder, “Is neocortex essentially multisensory?,” Review TRENDS in Cognitive Sciences Vol.10 No.6, pp.278-285, June 2006.

[29] M.S. Beauchamp, “See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex,” Curr Opin Neurobiol, 15, pp.1–9, 2005.

[30] I.R. Olson, J.C. Gatenby and J.C. Gore, “A comparison of bound and unbound audio–visual information processing in the human cerebral cortex,” Cognitive Brain Research 14, pp.129–138, 2002.

[31] H. Kafaligonul and G.R. Stoner, “Auditory modulation of visual apparent motion with short spatial and temporal intervals,” Journal of Vision 10(12), 31, pp.1–13, 2010.

[32] E. Freeman and J. Driver, “Direction of visual apparent motion driven solely by timing of a static sound,” Current Biology, 18, pp.1262–1266, 2008.

[33] I.R. Olson, J.C. Gatenby and J.C. Gore, “A comparison of bound and unbound audio–visual information processing in the human cerebral cortex,” Cognitive Brain Research 14, pp.129–138, 2002.

[34] G.A. Calvert, et al, “Response amplification in sensory-specific cortices during crossmodal binding,” Neuroreport, 10, pp.2619–2623, 1999.

[35] R.T. Born and D.C. Bradley, “Structure and function of visual area MT,” Annual Review of Neuroscience, 28, pp.157–189, 2005.

[36] C.C. Pack and R.T. Born, “Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain,” Nature 409(6823), pp.1040–1042, 2001.

[37] W. Bair and C. Koch, “Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey,” Neural Computation, 8, pp.1185–1202, 1996.

[38] G.T. Buracas, A.M. Zador, M.R. DeWeese and T.D. Albright, “Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex,” Neuron, 20, pp.959–969, 1998.

[39] K. Jantzen, F. Steinberg and J. Kelso, “Functional MRI reveals the existence of modality and coordination-dependent timing networks,” Neuroimage, 25, pp.1031–1042, 2005.

[40] D. Bueti, B. Bahrami and V. Walsh, “Sensory and association cortex in time perception,” Journal of Cognitive Neuroscience, 20, pp.1054–1062, 2008.

[41] L. Scheef, et al, “Multimodal motion processing in area V5/MT: Evidence from an artificial class of audio visual events,” Brain Research, 1252, pp.94–104, 2009.

[42] M. Saenz, L.B. Lewis, A.G. Huth, I. Fine and C. Koch, “Visual Motion Area MT/V5 Responds to Auditory Motion in Human Sight-Recovery Subjects,” The Journal of Neuroscience, 28, pp.5141–5148, 2008.

[43] A. Alink, W. Singer and L. Muckli, “Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex,” Journal of Neuroscience, 28, pp.2690–2697, 2008.

[44] V.A.F. Lamme and P.R. Roelfsema, “The distinct modes of vision offered by feedforward and recurrent processing,” Trends Neurosci. 23, pp.571–579, 2000.