論文 / 著書情報
Article / Book Information

題目(和文)	同時性知覚に対する運動情報の影響
Title(English)	The Effect of Motion Information on Simultaneity Perception
著者(和文)	権 眞煥
Author(English)	Jinhwan Kwon
出典(和文)	学位:博士（理学）, 学位授与機関:東京工業大学, 報告番号:甲第9961号, 授与年月日:2015年9月25日, 学位の種別:課程博士, 審査員:三宅　美博,中村　清彦,寺野　隆雄,新田　克己,宮下　英三
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第9961号, Conferred date:2015/9/25, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

The Effect of Motion Information on Simultaneity Perception

Doctoral Dissertation

Jinhwan Kwon

Department of Computational Intelligence and Systems Science
Interdisciplinary Graduate School of Science and Engineering
Tokyo Institute of Technology

Abstract

Motion perception is a fundamental tool for interaction with a dynamic environment. When considering real-time interaction, temporal information of moving object is a significant factor for smooth interaction. In particular, simultaneity perception is an important key for flexible human behavior with the dynamic environment. In other words, real time interaction with the dynamic environment is generated by the relationship between motion perception and simultaneity perception. This dissertation addresses the influence of visual motion information on simultaneity perception. The main aims of this dissertation were (1) to examine how visual apparent motion affects audiovisual simultaneity perception, (2) to investigate what mechanisms contribute to the findings that visual apparent motion affects simultaneity perception, (3) to examine the influence of visual motion on simultaneity perception in human communication. These aspects are described in more detail in Chapter 1.

In Chapter 2, the purpose was to investigate how visual apparent motion affects audiovisual simultaneity perception. Two types of temporal order judgment (TOJ) tasks were examined to confirm the effect of visual apparent motion. Participants conducted audiovisual TOJ tasks in the apparent motion condition with two flashes, and in the normal condition with a single flash, which is the conventional condition of a TOJ task. The results of the experiments showed that point of subjective simultaneity (PSS) in the apparent motion condition was shifted toward a sound-lead stimulus, which is closer to physical simultaneity (i.e., zero) and just noticeable difference (JND) in the apparent motion condition was smaller than that in the normal condition. This means that audiovisual simultaneity perception is closer to physical simultaneity from the PSS's result and showed greater temporal resolution from the JND's result during apparent motion perception. This results suggest that visual apparent motion contributes to very precise perceptions of temporal simultaneity in audiovisual integration.

In Chapter 3, the goal was to investigate what mechanisms contribute to the finding obtained in Chapter 2. Three possible mechanisms were considered and they were examined in three sections, respectively. In Section 1, the purpose was to examine the influence of amount of visual stimulation as the first possible mechanism, because it remained unclear whether the results obtained in Chapter 2 were influenced by

differences in the amount of visual stimulation. As a result, the PSS and JND obtained in the apparent motion condition differed from those obtained in the successive condition as non-motion perception, which included the same amount of visual stimulation of the apparent motion condition. This means that the amount of visual stimulation made no difference in the apparent motion condition.

In Section 2, the purpose was to investigate the influence of prediction (a higher-order brain function) as the second possible mechanism, because apparent motion was produced by a constant interval between two flashes. Therefore, it was necessary to eliminate the influence of prediction by randomizing the intervals between the two visual stimuli. As a result, the results obtained in Section 2 did not differ from those obtained in the apparent motion condition in Chapter 2. This means that apparent motion was equivalently processed regardless of prediction.

In Section 3, the purpose was to confirm whether motion binding property as the third possible mechanism influences audiovisual simultaneity perception. Visual apparent motion is achieved by binding property that stimuli are perceived as a moving object with spatiotemporal continuity, and a single bounded object. Therefore, there is a need to examine the influence of motion binding property. As a result, the PSS shifted toward a sound-lead stimulus and especially became closer to physical simultaneity in the motion binding condition, and the JND in the motion binding condition was smaller than those in the non-motion binding condition. This shows that motion binding property shifted simultaneity toward physical simultaneity, and reduces the temporal window in audiovisual integration. Therefore, it is revealed that motion binding property contributes to the accurate perceptions of temporal events in audiovisual processing.

In Chapter 4, the purpose was to investigate the influence of visual motion on simultaneity perception in human communication. To confirm the influence of visual motion, body movement synchronization is detected in face-to-face (visual interaction) and remote communication (unidirectional visual perception) settings using a new definition through phase difference because body movement synchronization is achieved by simultaneity perception. As a result, although the mean phase differences in head nods did not differ significantly between the face-to-face communication and remote communication conditions, there were significant differences in the densities, standard deviations and kurtoses in the phase difference distributions between the two conditions. This means that visual interaction resulted in higher synchronization activity and strength by accurate simultaneity perception.

This dissertation addresses the influence of visual motion information on simultaneity perception. The findings showed that visual motion information contributes to higher accuracy on simultaneity perceptions of temporal events in the dynamic environment. Therefore, motion perception allows flexible interaction with the dynamic environment in real time.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Simultaneity Perception in a Dynamic Environment

We interact with a dynamic environment by perceiving and interpreting motion, which has become the most important information in our daily life. Motion determines the perceptual frame in the dynamic environment. The physical phenomenon ‘motion’ is defined as the spatiotemporal change of an object. In human beings, motion can be broadly classified into two categories, namely ‘motion perception’ and ‘body motion’. Motion perception means that we perceive and recognize the movement of the object from a dynamic environment. Body motion means that we interact with the dynamic environment through action. We interact with the dynamic environment in real time by using motion perception and body motion. In particular, motion perception is a fundamental tool for interaction with the dynamic environment in real time and temporal information of motion is very important for smooth interaction with the environment.

Interestingly, it has been reported that there exists a temporal difference between the presentation of external sensory stimuli and the perception of stimuli (Spence et al., 2001; Spence et al., 2010). External sensory stimuli are perceived with perceptual latency due to transmission time through air and neural transmission time (Vroomen and Keetels, 2010). However, we interact with motion smoothly in real time in a dynamical environment. In order to interact with the environment smoothly, the relation between motion perception and temporal perception is important and especially ‘simultaneity perception’ in temporal perception plays significant role in interaction with the environment. Simultaneity perception enables to perceive multiple sensory information as a single event and interact with the environment at the same time. In other words, we can interact with the environment in ‘real time’ by simultaneity perception. Real time interaction with the dynamic environment is generated by the relationship between motion perception and simultaneity perception. Therefore, there is a need to focus on the relationship between motion perception and simultaneity perception.

1.2 Visual Motion Perception

We perceive motion through the multiple senses in the dynamic environment to translate continuously motion information from the external world. It is possible to include visual, auditory, tactile senses as typical senses to perceive motion information. In particular, visual motion perception is most deeply involved in the processing of motion information since vision is dominantly characterized by the spatiotemporal processing of motion information (Burr and Thompson, 2011; Adelson and Bergen, 1985; Krekelberg, 2008). There are two motion perception systems, which are first-order and second-order motion perception systems (Nishida, 2011; Vaina and Soloviev, 2004). The first-order motion system is defined as pure motion perception by the movement of luminance and color attributes. On the other hand, the second-order motion system is defined as the movement of high-level features such as contrast, texture and flicker.

In physiological aspect, visual signals received by retina are transmitted to primary visual cortex as lateral geniculate nucleus (LGN) and V1 (Van Essen, et al., 1992; Felleman and Van Essen, 1991; Hilgetag, et al., 1996; Casagrande and Norton, 1991; Merigan and Maunsell, 1993; Stone, et al., 1979). The visual signals are integrated and interpreted by hierarchical and parallel pathways, which are divided into two main pathways of dorsal and ventral stream (Goodale and Westwood, 2004; Fang and He, 2005; Martin, 1992; Nassi and Callaway, 2009; Callaway, 2005; Hendry and Reid, 2000; see Figure 1.1).

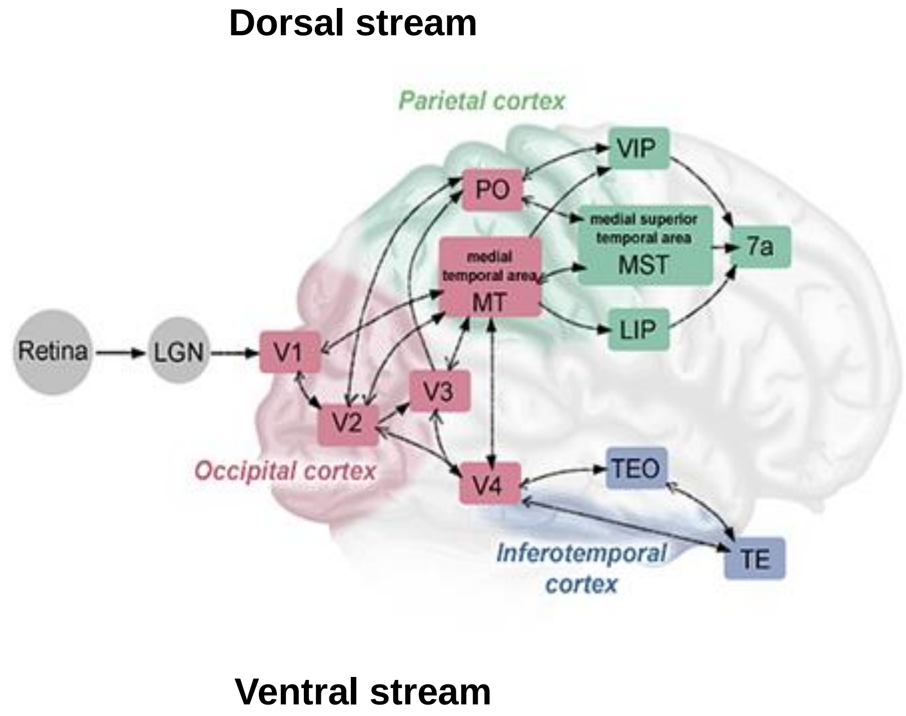


Figure 1.1. Schematic representation of the two streams of visual processing in human cerebral cortex (Goodale and Westwood, 2004). The retina sends projections to the dorsal part of the lateral geniculate nucleus (LGN) in the thalamus, which projects in turn to primary visual cortex (V1). Within the cerebral cortex, the ventral stream (red) arises from early visual areas (V1) and projects to regions in the occipito-temporal cortex. The dorsal stream (blue) also arises from early visual areas but projects instead to the posterior parietal cortex. The posterior parietal cortex also receives visual input from the superior colliculus through the pulvinar.

The dorsal stream is related to spatial characteristics, detecting movement and guiding actions, whereas the ventral stream is involved in object recognition and identification such as shape, color and depth. In particular, the dorsal stream plays an important role in visual motion processing and area MT (middle temporal/ V5) in the dorsal stream is established as a key substrate of visual motion processing. Area MT has a function which comprehensively processes the whole motion information of objects by integrating the information of local motion (Pack and Born, 2001). Area MT is well studied spatial and temporal tuning properties and neurons tuned to speed are rare in area V1 but are reported in area MT (Perrone & Thiele, 2001; Priebe, Lisberger, & Movshon, 2006). Area MT has been found to be activated when subjects engage in visual timing tasks such as the timing of rhythmic visual stimuli (Jantzen, Steinberg, & Kelso, 2005) and estimating the time of visual interception (Bosco, Carrozzo, & Lacquaniti, 2008). Bueti et al. (2008) found that area MT impaired the ability of human subjects to discriminate short temporal intervals (i.e. of hundreds of milliseconds).

Interestingly, we can experience the motion through not only continuous signal but also a discrete sequence of events (Figure 1.2). This property of motion perception on human perceptual system is used for various display systems such as film and television (Wu and Rao, 2005; Watson, et al., 1983; Watson, et al. 1986). For example, a film is presented by a sequence of static views, and we usually perceive the film with motion continuously and smoothly. In other words, although such display systems deliver a discrete sequence of static views, we cannot avoid perceiving it as continuous moving images under a specific condition (e.g., 24 fps in films, 30 fps in television), so-called “window of visibility” (Watson, et al., 1983; Watson, et al. 1986). In particular, it is reported that a unit of motion perception is characterized by discrete intervals, which is sampling time of 20 ms. Therefore, the motion perception on human perceptual system relies on this 20 ms sampling time, even real motion (Burt and Sperling, 1981). On the other hand, many researchers have reported that temporal resolution of motion perception is a range of 20-150ms on human perceptual system (Watson, 1985; Braddick, 1980; Larsen, 1983; Lamme, 2000) This property of motion perception is called ‘apparent motion’ (Watson, et al., 1983; Watson, et al. 1986, Burt and Sperling, 1981).

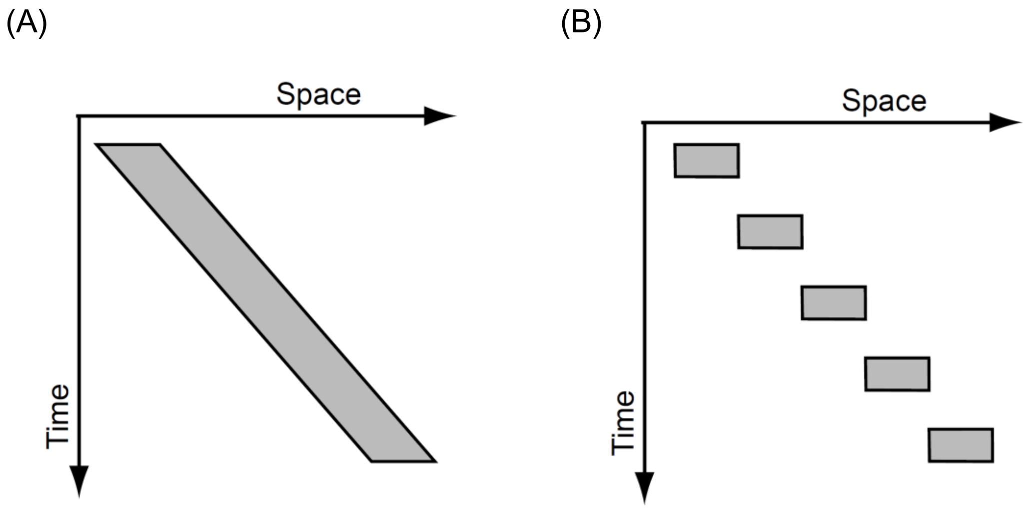


Figure 1.2. Motion as space-time orientation (Krekelberg, 2008). Contrast distribution of a sampled version of the moving line. Points indicate the times and positions at which the sample presented. (A) The distribution of a bar shows smooth and continuous motion over space and time. When a bar moves to the right over time, the motion is perceived clearly. (B) The distribution of the bar indicates sampled and discrete elements. When the bar is presented as discrete elements as jump from one place to the next (apparent motion), the motion is still clearly visible as smooth and continuous movement.

1.3 Apparent Motion

As mentioned above, visual motion perception are determined by temporal and spatial characteristics on human perceptual system. Especially, even despite two discrete stimuli, motion is perceived by the appropriate spatiotemporal interval (Ramachandran and Anstis, 1986; Anstis and Mackay, 1980; Larsen et al., 1983; Braddick, et al., 1980; see Figure 1.3). This shows that there exist some perceptual frames for perceiving continuous motion. Visual apparent motion is an optical phenomenon whereby motion appears to occur at a certain spatiotemporal interval, although two discrete stimuli are used (Ramachandran and Anstis, 1986; Anstis and Mackay, 1980; Larsen et al., 1983; Braddick, et al., 1980; Grossberg and Rudd, 1992; Dawson, 1991). In particular, visual apparent motion is systematically affected by the temporal interval between two stimuli. When the temporal interval is too short, the stimuli are perceived as simultaneous, whereas when the temporal interval exceeds a certain interval, the two stimuli are perceived as successive. Therefore, if the temporal interval between two stimuli is too short or too long, motion is not perceived. For example, many researchers have reported that two visual stimuli are perceived as a continuous motion when the interstimulus onset interval (ISOI) of the visual stimuli is within a range of 50 to 150 ms. Conversely, the visual stimuli are perceived as simultaneous within an ISOI of 40 ms and the visual stimuli are perceived as successive beyond an ISOI of 300 ms (Dawson, 1991; Getzmann, 2007; Harrar et al., 2008; Strybel, et al., 1990; Briggs and Perrott, 1972). Of particular note is that apparent motion is a fundamental unit of visual motion in finite time.

Visual apparent motion is divided into short- and long-range processes (Ramachandran and Anstis, 1986; Anstis and Mackay, 1980; Larsen et al., 1983; Braddick, et al., 1980; Dawson, 1991). Short-range process in visual apparent motion, called low-level process, is characterized by short spatial displacements (up to 30″ or less of visual angle) and brief temporal intervals (less than 100ms of inter-stimulus intervals). Also, long-range apparent motion, called high-level process, is known to be detected in much larger spatial displacements (several degrees and up

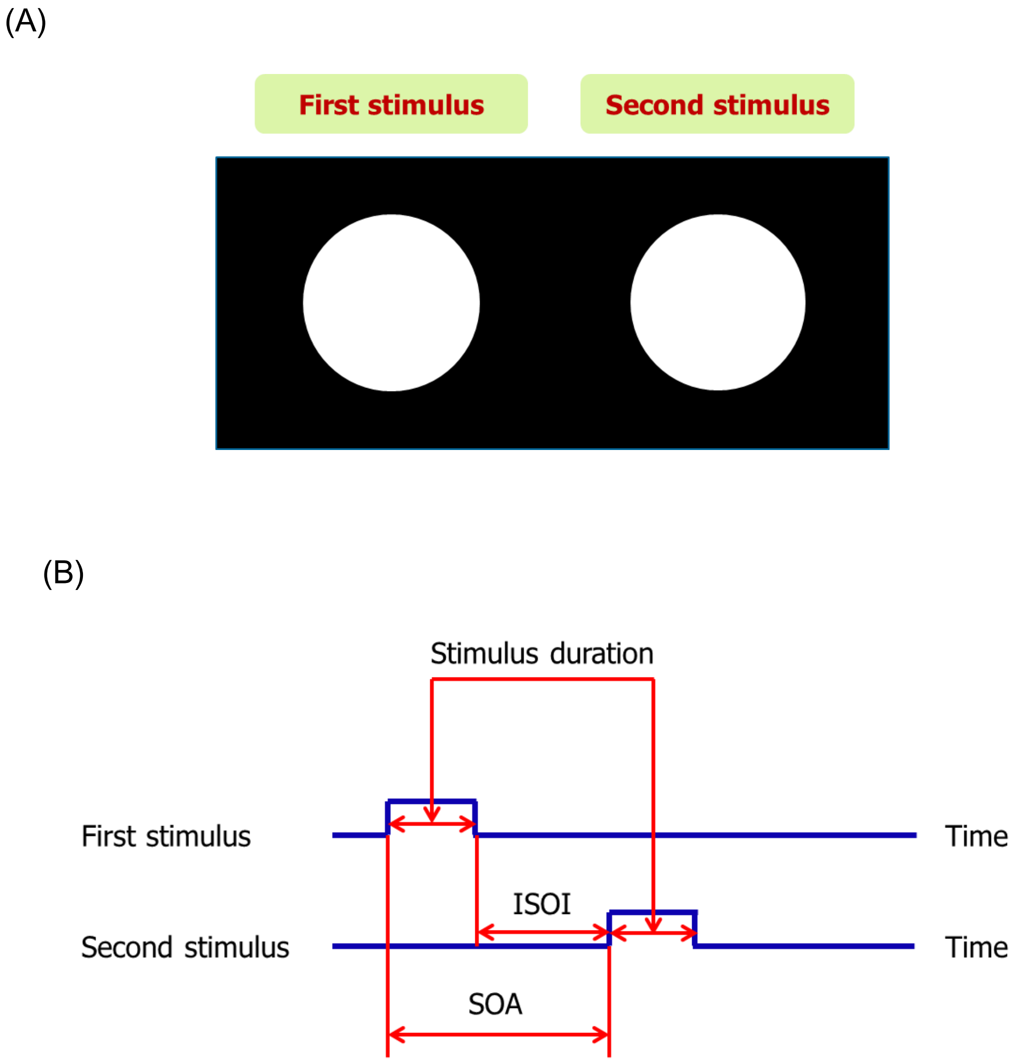


Figure 1.3. Apparent motion and temporal factors. (A) shows the structure of visual apparent motion. (B) shows temporal factors in perception of visual apparent motion. Two visual stimuli are perceived as a continuous motion when the interstimulus onset interval (ISOI) between the two visual stimuli is within a range of 50 to 150 ms. Conversely, the visual stimuli are perceived as successive beyond an ISOI of 300 ms (Dawson, 1991; Getzmann, 2007; Harrar et al., 2008; Strybel, et al., 1990; Briggs and Perrott, 1972).

to 15° of visual angle) and longer temporal intervals (inter-stimulus intervals of more than 100ms and up to 300ms). The two processes have been studied in a wide range of fields from psychophysics to brain functions (Grossberg and Rudd, 1992; Dawson, 1991). In particular, psychophysical aspects have been studied to elucidate the perceptual or cognitive characteristics of apparent motion, and the temporal factors are considered to be important for apparent motion perception.

In the physiological aspect, the response to apparent motion is typically equivalent to the response to real motion in the physiological mechanisms by passing through the dorsal stream associated with motion processing (Newsome, et al., 1986; Shimono, et al., 2012; Manning, et al., 1988; Shadlen, et al., 1993). Moreover, some researchers reported a possibility that higher cortical areas are implicated in the processing of apparent motion (Newsome, et al., 1986).

1.4 Motion and Temporal Perception

In unisensory processing, it has been reported that motion is perceived faster than non-motion information (Schlag and Schlag-Rey, 2002; Hikosaka, et al., 1993; Nijhawan, 1994). For example, there is a robust illusory phenomenon called flash-lag effect, which means that motion is perceived faster than a single flash (Nijhawan, 1994; Nijhawan, 2002; Vroomen and Gelder, 2004). When a moving target and a static flash are aligned and when they appeared at the same location, the moving target is perceived as shifted more slightly in the direction of motion relative to the flash. In other words, the flash is processed more slowly than the moving target. However, Brenner and Smeets (2000) pointed out the problem of flash-lag experiment. Brenner and Smeets (2000) claimed that a moment for spatial alignment of moving object and a flash, especially moving target’s location, can only be referred after the flash’s presentation. In other words, even if the moving object and the flash were presented with the same physical location, the moment for spatial alignment was determined after the flash was perceived.

Besides, Vroomen and Gelder (2004) examined whether a sound presented with a flash at the same time has an effect on the flash-lag effect. Participants judged the position of a flash presented at various timings relative to a moving bar that moved laterally from the left side to the right side with constant speed. They compared a synchronized sound condition and silent condition. Synchronized sound condition shows that a sound was synchronized with a flash and silent condition indicates that a sound was not presented with a flash. The results showed that when the sound was present with a flash, the flash-lag effect were reduced relative to the silent condition. Consequently, they reported that a sound can enhance temporal processing of a flash. However, there is a possibility that participants might judge the position of a flash by the single sound as pointed out by Brenner and Smeets (2000). In other words, the single sound might determine the timing of the single flash because the single sound is perceived faster than the single flash. Furthermore, it is difficult to understand whether the single sound had an effect on the motion or the flash. Therefore, it is necessary to examine the temporal characteristics by dividing motion information and non-motion information, including single flash. This study is focused on the temporal perception during the observation of motion and non-motion information in multisensory integration.

1.5 Temporal Integration in Multisensory Processing

1.5.1 Unity Assumption

When a single auditory and visual stimuli occurred from external world, we can perceive them as a same unitary event. On the other hand, we can perceive them as the two separate events without integration as well. The binding or separation of these unimodal stimuli is important to process the external information from the environment. In particular, the binding of multisensory stimuli is known as “unity assumption,” which means that we perceive the multisensory information as a unitary information (Bedford, 2001; Radeau and Bertelson, 1977; Bertelson and

Aschersleben, 1998; Vatakis and Spence, 2007; Welch, 1999; Welch and Warren, 1980). Researches on the unity assumption have been demonstrated in various dimensions such as time, space, semantics and pattern on the basis of the consistency of the information and perceptual binding (Welch and Warren, 1980; Laurienti, et al, 2004; Vatakis and Spence, 2006; Vatakis and Spence, 2007; Sanabria, et al, 2004; Spence, et al, 2007; Guski and Troje, 2003). In particular, spatiotemporal origin of different modalities is important factor to bind them into a single perceptual object or event (Jack and Thurlow, 1973; Radeau & Bertelson, 1977; Slutsky and Recanzone, 2001; Bertelson and Aschersleben, 2003; Morein-Zamir, et al., 2003). The crossmodal binding of multiple sensory information as a single object or event has been studied using a simple set of stimuli, such as a combination of a single sound and a single flash, and more complex stimuli such as audiovisual speech (Easton and Basala, 1982; Wassenhove, et al., 2007; Vatakis and Spence, 2006; Vatakis and Spence, 2007; Walker, et al, 1995; Green and Gerdeman, 1995; Green, et al, 1991).

1.5.2 Space Dimension

The multisensory binding as unity assumption has been demonstrated focusing on ‘capture’ phenomenon, which means that one or more sensory information is often captured by another sensory information. In other word, when two or more sensory information is transmitted with different speeds, multisensory perception is usually biased toward one modality, so called sensory or modality dominance. For example, many studies have reported ventriloquism effect with regard to capture phenomenon in space dimension. The spatial ventriloquism effect shows that when auditory and visual stimuli are presented at the same time, auditory stimuli are typically captured toward location of visual stimuli (Bermant and Welch 1976; Bertelson and Radeau 1981; Jack and Thurlow, 1973; Radeau, 1994; Radeau and Bertelson, 1977; Hairston, et al, 2003; Slutsky and Recanzone, 2001). That is, although the location between audiovisual stimuli is different, the perceived location of auditory stimuli shifts toward the position of visual stimuli and the audiovisual information is integrated as a same event. Therefore, although the positions of the screen and speakers in a movie theater are different we can catch the audiovisual stimuli as a single unit of information. The spatial ventriloquism effect, vision dominates audition in space dimension, is a well-known phenomenon in multisensory integration (Radeau 1994; Bertelson and Aschersleben 1998).

1.5.3 Time Dimension

This sensory capture phenomenon has also been demonstrated in the time dimension. Previous studies have reported that the temporal integration of multisensory information depends on the combination of sensory information and it can be changed according to a variety of factors such as stimulus intensity, prediction and attention (Allik and Kreegipuu, 1998; Jakowśki and Verleger, 2000; Mattes and Ulrich, 1998; Schneider and Bavelier, 2003). In particular, many studies have shown that audition dominates vision on multisensory processing in the time dimension (Gebhard and Mowbray, 1959; Fendrich and Corballis, 2001; Bertelson and Aschersleben, 2003; Morein-Zamir, et al., 2003; Freeman and Driver, 2008; Recanzone, 2003; Vroomen and de Gelder, 2004). Auditory dominance or auditory driving, whereby audition captures vision, is a widely known phenomenon (Gebhard and Mowbray, 1959; Fendrich and Corballis, 2001). For example, Fendrich and Corballis (2001) reported that a flash was perceived significantly earlier when preceded by an auditory stimulus. On the other hand, a flash was seen significantly later when followed by an auditory stimulus. In addition, temporal ventriloquism effect has shown that temporal perception of visual stimuli is pulled into that of auditory stimuli (Fendrich and Corballis, 2001; Morein-Zamir, et al., 2003; Scheier, et al, 1999; Vroomen and Keetels, 2006). For example, Morein-Zamir S, et al., (2003) reported that the temporal ventriloquism effect condition shifted the temporal window of integration toward smaller values than those of the baseline. The temporal ventriloquism effect condition shows that one auditory stimulus was presented shortly before the first light, and the other after the second light and the baseline condition shows that each auditory stimulus was presented with a flash at the same time (i.e., each pair of an auditory stimulus and a visual stimulus was presented at the same time). This shows that sounds attracted the temporal perception of lights by improving the visual temporal resolution. Therefore, the phenomenon, which audition dominates vision in time dimension, has been considered as a mainstream in multisensory temporal perception, especially with higher temporal resolution and acuity on temporal perception.

1.5.4 Physical and Physiological Aspects

When we interact with the environment, the multisensory signals reached the sensory receptors respectively and are perceived with relative time. It is reported that the perceived order of the sensory stimuli is different in the brain because the physical transmission time through air and neural transmission time are different in kind of sensory stimuli (Vroomen and Keetels, 2010; Spence et al., 2001; Spence et al., 2010; Spence and Squire, 2003; King, 2005). Physical transmission time of light is approximately 300,000,000 m/s and physical transmission time of sound is approximately 330 m/s (see Table 1-1). Therefore, physical transmission time of sound is slower than that of light. Light and sound reach ears and eyes, and then the auditory and visual signals are transmitted to the brain by neural process from the sensory receptors. The neural processing time differs between light and sound, and the neural processing time of light is typically slower than that of sound (approximately 50 vs. 10 ms, respectively). This shows the temporal disparity by relative speed from physical and neural transmission time between light and sound. Although the multisensory information is processed by the temporal disparity we perceive and integrate the multisensory information as simultaneous event. However, the mechanism about how the temporal simultaneity in multisensory integration is maintained is not clear. It is very interesting point that whether the brain cannot discriminate the small lag between multisensory signals or whether the brain flexibly controls the multisensory signals to perceive the synchrony or asynchrony (See Figure 1.4).

Table 1-1. Physical transmission time through air and neural transmission time between light and sound.

Physical transmission times	
Light	**Sound**
300,000,000 m/s	330 m/s
Neural transmission times	
Light	**Sound**
50ms	10ms

Physical transmission time of light is approximately 300,000,000 m/s and physical transmission time of sound is approximately 330 m/s. Therefore, physical transmission time of sound is slower than that of light. The neural processing time differs between light and sound, and the neural processing time of light is typically slower than that of sound (approximately 50 vs. 10 ms, respectively). This shows the temporal disparity by relative speed from physical and neural transmission time between light and sound.

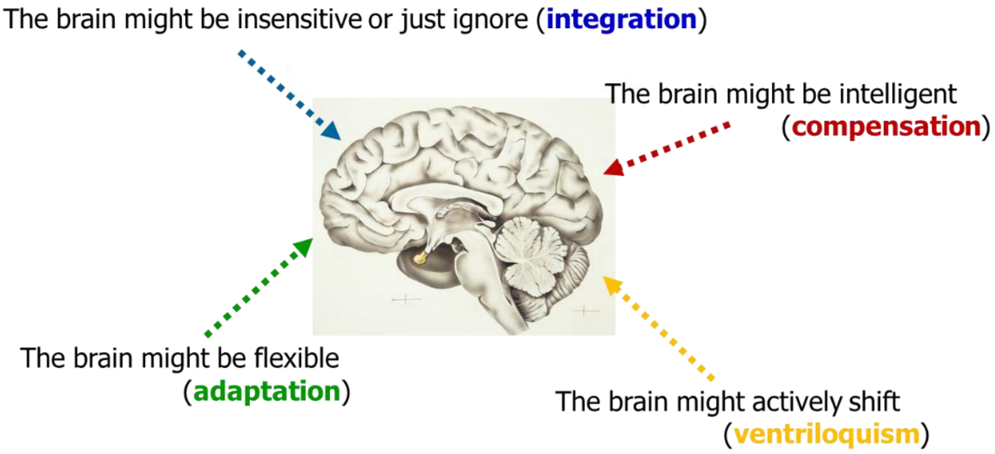


Figure 1.4. How intersensory synchrony might be maintained? (Vroomen and Keetels, 2010)

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 (Stein and Stanford, 2008; Ghazanfar and Schroeder, 2006; Beauchamp, 2005; see Figure 1.5). However, in resent year, several functional imaging studies have also showed the evidences that the audiovisual integration is related to area MT engaged in visual motion processing (Olson, et al., 2002; Calvert, et al., 1999). Furthermore, some researchers suggested that audiovisual stimulation is integrated in Area MT under the dorsal stream implicated in motion processing (Scheef, et al., 2009; Saenz, et al., 2008; Alink, et al., 2008). These findings collide with the brain areas, especially superior temporal sulcus (STS), known as traditional audiovisual integration. However, some researchers have suggested that the mechanism of multisensory processing by motion information diffed from that of multisensory processing by non-motion information (Kafaligonul and Stoner, 2010, Scheef, et al., 2009; Alink, et al., 2008). Therefore, there is a need to quantitatively investigate the influence of motion and non-motion information on the temporal integration of multisensory perception.

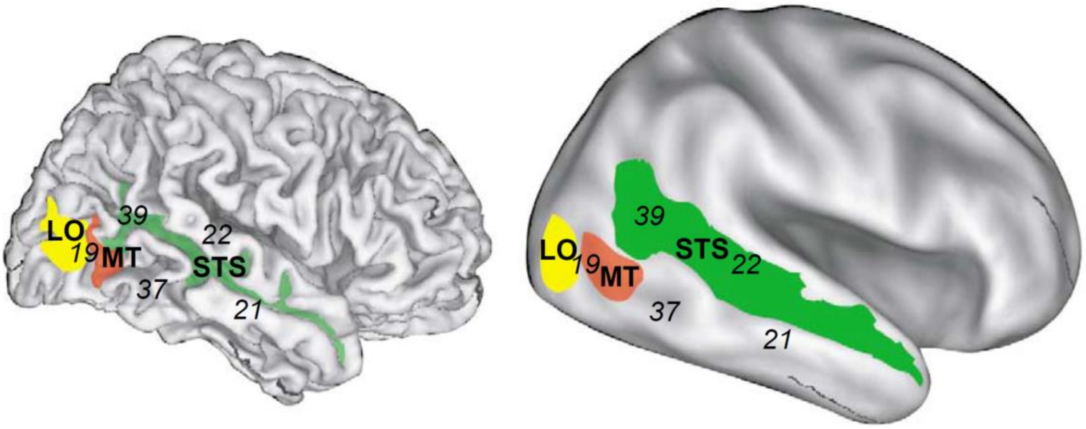


Figure 1.5. The locations of the three multisensory regions (Beauchamp, 2005). The locations of the three multisensory regions described in the review shown on a lateral view of a folded (left) and inflated (right) right hemisphere: area LO is yellow, area MT is red and STS is green. Numbers indicate approximate centers of nearby Brodmann areas. The region labeled LO corresponds retinotopically to dorsal V4 and is a subset of a much larger band of cortex (extending superiorly and inferiorly) that responds preferentially to real visual images versus scrambled controls.

1.6 Simultaneity Perception

1.6.1 Temporal Order Judgment Task

With respect to multisensory integration, the important question is about how multiple senses are integrated in the time dimension. We can perceive multiple sensory information as a single event. It is made possible by simultaneity perception. Therefore, simultaneity perception is an important topic in the study of multisensory integration (Pöppel, 1988).

To confirm the characteristics of simultaneity perception, it is important to examine the temporal characteristics between external stimuli and sensory perception. There has been remarkable progress in simultaneity judgment (SJ) and temporal order judgment (TOJ) tasks in psychophysical studies that investigate temporal factors in multisensory integration (Eagleman, 2008; Ivry and Schlerf, 2008; Grondin, 2010; Mauk and Buonomano, 2004). In particular, TOJ tasks are a traditional way to measure temporal characteristics between two or more senses. Specifically, TOJ task is a study to examine the temporal relationship between order of external stimuli and the temporal order perception in multisensory processing (Vroomen and Keetels, 2010; Spence et al., 2001; Spence et al., 2010). In this method, the point of subjective simultaneity (PSS) and just noticeable difference (JND) are two important parameters as the methods of measurement (Vroomen and Keetels, 2010; see Figure 1.6). The PSS represents the interval between the applications of stimuli to two senses at which both are perceived by the senses as occurring at the same time, which makes it possible to detect which sensory information was perceived early or late. The JND can be used as an indicator of temporal resolution in cross-modality, which discriminates the temporal asynchrony in multisensory processing.

In TOJ tasks using a set of audio and visual stimuli, many studies have demonstrated that hearing changes or attracts visual temporal perception, which is called ‘temporal ventriloquism effect’ (Bertelson and Aschersleben, 2003; Morein-Zamir, et al., 2003). For example, Morein-Zamir et al. (2003) investigated whether auditory events can alter the temporal perception of visual events using a visual

TOJ task. As a result, TOJ performance was improved with a smaller JND when one auditory stimulus was presented shortly before the first light and the another after the second light. Although many studies have reported that PSSs and JNDs depend on various factors in TOJ tasks such as spatial separation, frequency and predictive information (Vroomen and Keetels, 2010; Spence et al., 2001; Spence et al., 2010; Allik and Kreegipuu, 1998; Jakowśki and Verleger, 2000; Mattes and Ulrich, 1998; Schneider and Bavelier, 2003), it remains unclear how motion information influences audiovisual temporal perception. Furthermore, there are any doubts that audition changes the visual perception in temporal domain. Most studies have used a single static stimulation so far. In addition, although some researchers have reported some outcomes by various factors, they do not take into account the presence or absence of movement in the factors. In case of single or static stimuli, auditory stimulation is likely to be stronger, but the changes by motion might have a greater influence on vision. In recent years, some researchers have raised the possibility that there exist different mechanisms by motion information in multisensory processing with the growing interest in the multisensory property of motion (Kafaligonul and Stoner, 2010). Therefore, we focused on the relationship between visual motion and simultaneity perception through apparent motion and TOJ task.

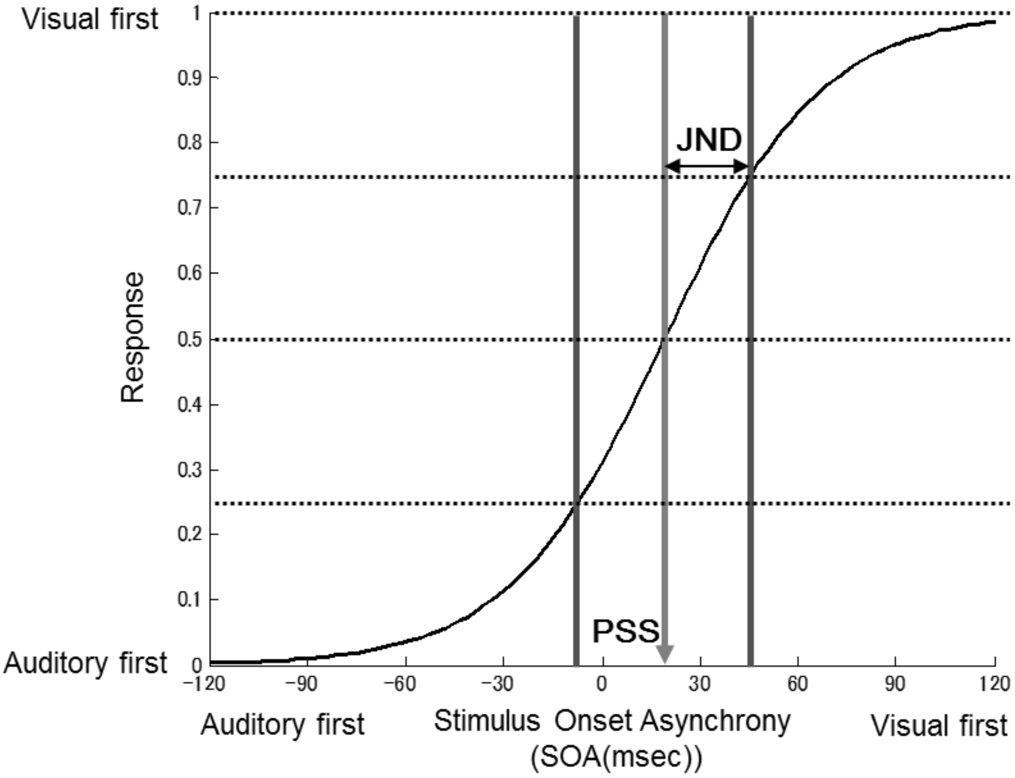


Figure 1.6. Logistic regression curve for data analysis of an audio–visual TOJ task. The point of subjective simultaneity (PSS) represents a specific interval between the applications of two sensory stimuli at which both are perceived at the same time. The just noticeable difference (JND) represents temporal resolution for identifying the simultaneity.

1.7 Remaining Issues

This dissertation address the influence of visual motion information on simultaneity perception. Previous studies have reported that the temporal integration of multisensory information depends on the combination of sensory information and they can be changed according to a variety of factors such as stimulus intensity, prediction and attention (Allik and Kreegipuu, 1998; Jakowśki and Verleger, 2000; Mattes and Ulrich, 1998; Schneider and Bavelier, 2003). In particular, many studies have shown that audition dominates vision on multisensory processing in the time dimension (Gebhard and Mowbray, 1959; Fendrich and Corballis, 2001; Bertelson and Aschersleben, 2003; Morein-Zamir, et al., 2003; Freeman and Driver, 2008; Recanzone, 2003; Vroomen and de Gelder, 2004). With respect to simultaneity perception, many studies have also demonstrated that hearing changes or attracts visual temporal perception (Bertelson and Aschersleben, 2003; Morein-Zamir, et al., 2003). In addition, many studies find that simultaneity perception depend on a variety of multisensory information in TOJ tasks such as spatial separation, frequency and predictive information (Vroomen and Keetels, 2010; Spence et al., 2001; Spence et al., 2010; Allik and Kreegipuu, 1998; Jakowśki and Verleger, 2000; Mattes and Ulrich, 1998; Schneider and Bavelier, 2003). However, it remains unclear how motion information influences audiovisual temporal perception. Therefore, we focused on the relationship between visual apparent motion with respect to motion information and TOJ task with regard to simultaneity perception. The main aims of this dissertation were (1) to examine how visual apparent motion affects audiovisual simultaneity perception, (2) to investigate what mechanisms contribute to the finding that apparent motion affects simultaneity perception, (3) to examine the influence of visual motion on simultaneity perception in human communication.

1.8 Overview of This Dissertation

The main purpose of this dissertation was to investigate the effect of visual motion information on simultaneity perception. This dissertation consisted of six chapters.

In Chapter 2, the purpose was to examine how visual apparent motion affects audiovisual simultaneity perception. Two types of TOJ tasks was examined to confirm whether visual apparent motion has an effect on an audiovisual TOJ task. Participants conducted audiovisual TOJ tasks in the apparent motion condition with two flashes, and in the normal condition with a single flash, which is the conventional condition of a TOJ task. The PSS and JND in the apparent motion condition were compared with those in the normal condition. Through Chapter 2, the effect of visual apparent motion on audiovisual simultaneity perception was confirmed.

In Chapter 3, the goal was to investigate what mechanisms contribute to the findings of Chapter 2 that visual apparent motion affects audiovisual simultaneity perception. Three possible mechanisms were considered in the Chapter 3, which consisted of three sections for examining the possible mechanisms. In Section 1, the purpose was to examine the influence of amount of visual stimulation by eliminating the influence of the amount of flash stimulation, because it remained unclear whether the results were influenced by differences in the amount of visual stimulation between the apparent motion condition with two flashes and the normal condition with a single flash. In Section 2, the purpose was to investigate the influence of prediction as a higher-order brain function, because apparent motion was produced by a constant interval between two flashes. In Section 3, the purpose was to confirm whether motion binding property in visual perception influences audiovisual simultaneity with a brief sound. Through Chapter 3, the mechanisms in which visual apparent motion affects audiovisual simultaneity perception were confirmed.

In Chapter 4, the purpose was to investigate the influence of visual motion on simultaneity perception in human communication. In addition, Chapter 4 provides a definition of phase difference for detecting body movement synchronization in human communication. To confirm the influence of visual motion, we detected the body movement synchronization in direct face-to-face (bilateral perception of visual motion) and remote communication (unilateral perception of visual motion) settings using a new definition of phase difference. Through Chapter 4, the influence of visual motion on simultaneity perception in human communication and the effectiveness of newly-devised method for detecting nonverbal synchronization were confirmed.

Finally, the findings obtained throughout this dissertation are summarized in Chapter 5 and 6.

Chapter 2

The Effect of Motion on Simultaneity Perception

2.1 Introduction

This study addresses the influence of visual motion on the temporal integration of multisensory information from an environment. Visual motion information from a dynamic environment is an influential factor in temporal perception and especially in multisensory temporal perception. However, it is unclear how visual motion information influences the temporal integration of multisensory perceptions. With respect to motion information, this study is focused on visual apparent motion. Visual apparent motion is a phenomenon in which two flashes presented in sequence in different positions are perceived as continuous motion. The purpose of the Chapter 2 is to investigate how visual apparent motion affects audiovisual temporal perception. We examined two types of TOJ tasks, and we examined whether visual apparent motion has an effect on an audiovisual TOJ task. Participants conducted audiovisual TOJ tasks in the apparent motion condition with two flashes, and in the normal condition with a single flash, which is the conventional condition of a TOJ task.

2.2 Method

2.2.1 Task Design

We conducted audiovisual TOJ tasks in this experiment in which the spatial location and duration of stimuli were identical. In this chapter, we examined whether visual apparent motion had an effect on an audiovisual TOJ task. Participants performed the TOJ task in the apparent motion condition in which two flashes were presented with a stimulus onset asynchrony (SOA) of 137 ms (Harrar, et al., 2008), and in the normal condition with a single flash.

2.2.2 Participants

Eighteen participants (16 males and two females, with a mean age of 24.3 years) participated in this experiment. Participants were paid to take part in the experiments, and written informed consent was obtained. These experiments were approved by the ethics committee of the Tokyo Institute of Technology.

2.2.3 Apparatus and Stimuli

All TOJ task experiments were conducted in a dark and soundproof room (0.00–0.01 cd/m2 luminance). Figure 2.1 illustrates the setup for the experiments. Visual stimulation was provided by a 27-inch LCD display (Samsung S27A950D, Korea) with a screen resolution of 1920 × 1080 pixels and a refresh rate of 120 Hz. The display was operated from a PC workstation (Apple Mac Pro, 3.2 GHz Quad-Core Intel Xeon, ATI Radeon HD 5770 graphic card, 1 GB GDDR5 memory, Cupertino, CA, USA) placed in front of the participants. 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 of 3.2 cm in diameter on a black background. The luminance of the black background on the screen was 0.09 cd/m2, and that of the white disks was 74.8 cd/m2. The visual angle was 2.8° for the single stimulus and 5.6° for two visual stimuli. Sound stimuli were presented as monaural sounds (65 dB, 1,000 Hz) delivered via two speakers (MM-SPWD3BK, Sanwa Supply, Japan). The speakers were located on top of the screen. These visual and auditory stimuli were generated and operated with a computer program (Matlab and Psychtoolbox-3, MA, USA).

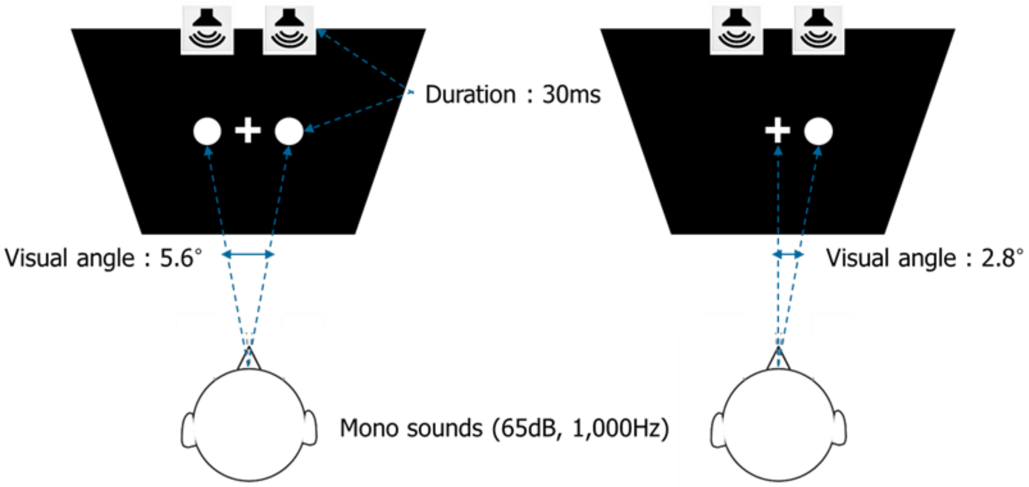


Figure 2.1. Schematic illustration of experimental setup. Figure 2.1 shows the setup for the experiments. Participants sit in front of a monitor with loudspeakers placed on top and a fixation point of a white cross displayed in the center. The visual angle was 2.8° for the single stimulus and 5.6° for two stimuli. Sound stimuli were presented via the two speakers.

2.2.4 Procedure

In the present experiment, the participants sat on a chair facing the stimulus, and a constant head position was maintained by means of the chin rest. The audiovisual TOJ tasks were performed over two sessions with visual stimuli in the apparent motion condition and in the normal condition. Figure 2.2 illustrate the procedure for this experiment. In the apparent motion condition (Figure 2.2), each trial began with display of the fixation cross for 1.5 s, followed by a dark blank screen for 800 ms. Next, one white circle for the first visual stimulus was displayed for 30 ms, and the second stimulus was presented with an SOA of 137 ms for 30 ms. To assess the temporal discrimination of a pair of auditory and visual stimuli, one brief sound (30 ms) was presented at various times relative to the second visual stimulus. The participants were instructed to complete a TOJ task between the second visual stimulus 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, after 500 ms, the participants 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 “light first” response was selected when the flash was ahead of the sound, and the “sound first” response was selected when the sound preceded the flash. In the normal condition (Figure 2.2), the first visual stimulus was not presented. That is, only the second stimulus in the apparent motion condition was shown in this session, and the other process was the same as that in the TOJ task in the apparent motion condition. The same method of evaluating the temporal discrimination between sound and flash with the same SOA values was then used as in the apparent motion condition. The present experiment consisted of 270 trials (2 visual conditions × 9 audiovisual SOAs × 15 repetitions) in counterbalanced order and the experiment was divided into 10 blocks of 27 trials (9 audiovisual SOAs × 3 repetitions). Including practice for each task, each experiment took approximately one and a half hours.

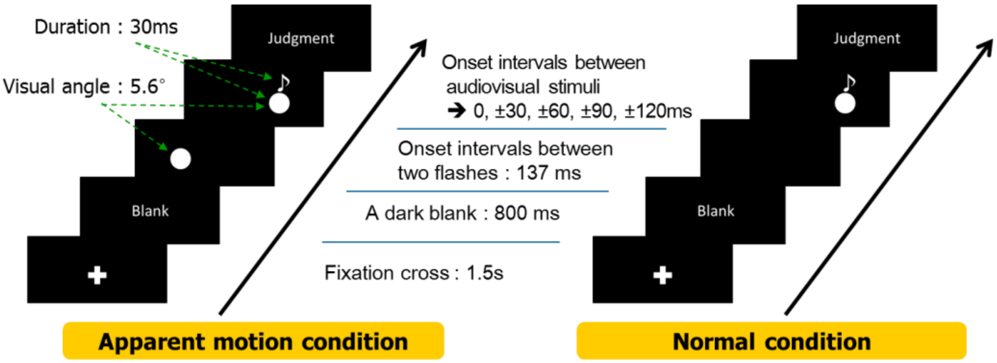


Figure 2.2. Schematic illustration of this experiment. Figure 2.2 shows the procedure for the present experiment. After the presentation of a fixation cross (1.5 s) and a blank screen (800 ms), two visual stimuli were presented with an SOA of 137 ms in apparent motion condition, or a single visual stimulus was presented in normal condition. Sounds were presented either before or after the visual stimuli at SOAs ranging from –120 to 120 ms at 30 ms intervals in random order.

Before starting each experiment, we examined whether the participants perceived motion between two flashes. In the present experiment, we showed the participants two flashes with an SOA of 137 ms 10 times, and participants were asked to evaluate whether their impression of the stimuli was of "continuous motion" or "successive stimuli."

2.2.5 Data Analysis

The ratio of answers indicating earlier presentation of an auditory stimulus was calculated for each SOA. We conducted logistic regressions using a generalized linear model with the ratio data from each experiment (Finney, 1952). The following equation was applied in the regression analysis:

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

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

$$JND = \frac{X_{75} - X_{25}}{2} = \beta \log 3, \tag{2}$$

where X_p represents the SOA with p percent of "visual first" responses.

As Figure 2.3 illustrates, psychometric curves were fitted to the distribution of the mean TOJ data for each condition. We determined the JND and PSS values for each participant using regression analysis (Equations (1) and (2)) and calculated mean and standard error values from the data.

2.3 Results

The results of two participants were excluded because they did not perceive continuous motion. Figure 2.3 shows the results of the experiment. Figure 2.3 (A) illustrates psychometric curves fitted to the distribution of the mean TOJ data for all participants in the apparent motion and normal conditions. Horizontal axis shows the temporal order of external stimuli and vertical axis shows the temporal order on perception. The horizontal axis shows the onset time of the auditory stimulus paired with a visual stimulus and the negative values of the SOAs in the horizontal axis indicate that an auditory stimulus preceded a visual stimulus. The vertical axis indicates the percentage of “visual first” responses to various stimulus onset asynchronies (SOAs) between audiovisual stimuli.

The Point of Subjective Simultaneity (PSS) is 50% crossover point of a distribution obtained by calculating the percentage of “visual first” responses to various stimulus onset asynchronies (SOAs) between stimuli. Zero up to 50 percent in the vertical axis shows the area of temporal order perception of “sound first”. On the other hand, 50 up to 100 percent in the vertical axis shows the area of temporal order perception of “light first”. Therefore, 50 percent crossover point in the vertical axis means the perceptual boundary that changes from sound first perception to light first perception. Consequently, the 50 percent crossover point is the SOA at which observers were maximally simultaneous (i.e., maximally unsure) about the temporal orders.

The just noticeable difference (JND) is obtained by a half the difference in SOAs between the 25% and 75% point. The JND shows the steepness of the graph, and represents the temporal resolution in cross-modality, which discriminates the temporal asynchrony in multisensory processing. In other words, JND indicates sensitivity to temporal asynchronies between multisensory information. Smaller JND means higher temporal resolution in cross-modality and higher sensitivity to temporal asynchronies.

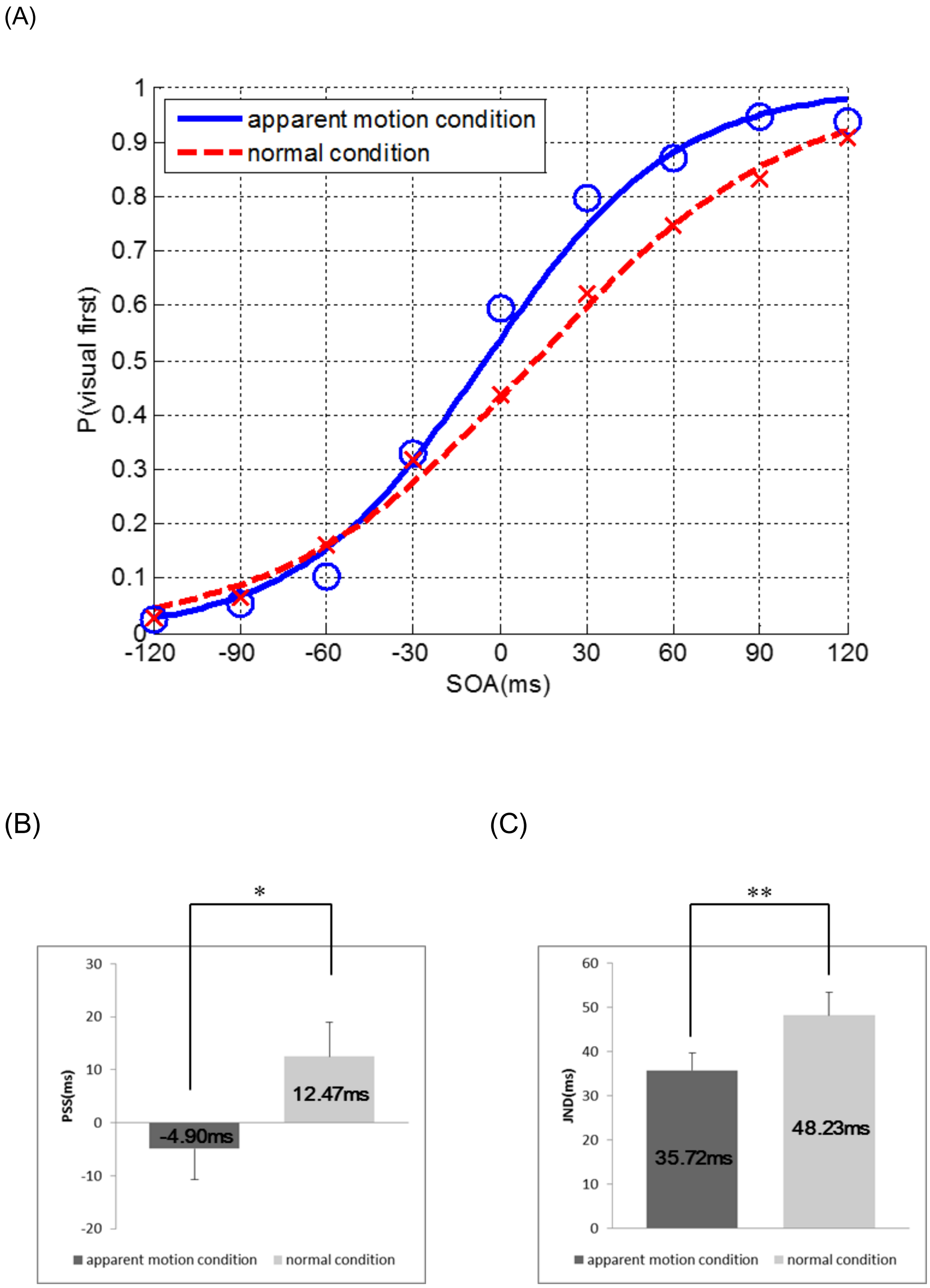


Figure 2.3. Results of the experiment. (A) shows psychometric curves fitted to the distribution of mean TOJ data for all participants in the apparent motion condition and in the normal condition. Horizontal axis shows the onset time of the

auditory stimulus paired with a visual stimulus. The negative values of the SOAs indicate that an auditory stimulus preceded a visual stimulus. (B) shows the mean PSSs in the apparent motion and normal conditions. Error bars represent the standard error of the means. (C) shows the mean JNDs in the apparent motion and normal conditions. Error bars represent the standard error of the means, *: p < .05, **: p < .01, for a paired t-test.

As shown in Figure 2.3 (B), the PSS in the normal condition had a positive value of 12.47 ms (SE = 6.45), but the PSS in the apparent motion condition shifted toward a sound-lead stimulus of –4.90 ms (SE = 5.84). This result indicates that a pair of audiovisual stimuli was perceived simultaneously when the auditory stimulus preceded the visual stimulus. A paired t-test of PSSs indicated a significant difference between the TOJ tasks in the apparent motion condition and those in the normal condition (t(15) = –2.33, p = 0.034, see Table A-1). In addition, the JND in the apparent motion condition was smaller than that in the normal condition (see Figure 2.3 (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 = 0.001, see Table A-2).

2.4 Discussion

The results of the experiment show that the PSS in the apparent motion condition was shifted toward a sound-lead stimulus, which differs from the PSS in the normal condition. Moreover, the JND in the apparent motion condition was smaller than that in the normal condition. Both conditions of the experiment had the same timing and spatial position to conduct the TOJ judgments. If physical transmission times through air and neural transmission times were considered it would be natural for auditory stimuli to be reached quickly in the brain (Vroomen and Keetels, 2010). However, the PSS in the apparent motion condition was shifted toward a sound-lead stimulus and the JND in the apparent motion condition was smaller than that in the normal condition. We use these results to clarify the influence of visual apparent motion on audiovisual simultaneity perception.

Visual apparent motion changes temporal simultaneity perception and improves temporal discrimination in audiovisual processing. With respect to temporal simultaneity, the experiment shows that the PSSs in the normal condition are similar to those of previous studies, which were usually shifted toward a visual-lead stimulus (Vroomen and Keetels, 2010; Mauk and Buonomano, 2004), whereas the PSSs in the apparent motion condition were shifted toward a sound-lead stimulus. Previous studies have been conducted using a simple set of stimuli, such as a combination of a single sound and a single flash in audiovisual simultaneity. For example, the temporal ventriloquism effect did not affect the baseline PSS. In the temporal ventriloquism effect condition, one auditory stimulus was presented shortly before the first light, and the other after the second light. In the baseline condition, each auditory stimulus was presented at the same time as a flash, (i.e., each pair of an auditory stimulus and a visual stimulus was presented at the same time) (Morein-Zamir S, et al., 2003) (see Table 2-1). It has also been reported that the PSS of audiovisual simultaneity perception usually shifts toward a visual-lead stimulus, so maximal simultaneity is perceived if light comes slightly before sound (Jakowski, et al., 1990; Lewald and Guski, 2003; Zampini, et al., 2005; Kanabus, et al., 2002; Kayser, et al., 2008; Wassenhove. et al., 2007; Zampini, et al., 2003).

Furthermore, several studies reported that the PSS changed under audiovisual simultaneity, in which the PSS became closer to physical simultaneity (i.e., zero). For instance, in audiovisual speech, the PSS of congruent audiovisual speech stimuli (the mean PSS of 23 ms) shifted closer to physical simultaneity than that of incongruent audiovisual speech under the McGurk effect (a mean PSS of 37 ms; see Table 2-1) (Wassenhove, et al., 2007). Moreover, Zampini et al. (2003) reported that the PSS was closer to physical simultaneity when audiovisual stimuli were presented in the same spatial position than when they were presented in different spatial positions (see Table 2-1) (Zampini, et al., 2003). This means that the same spatial positions of simple audiovisual stimuli and congruent utterance information in audiovisual speech shifted the PSS toward physical simultaneity in audiovisual integration, but the tendency toward visual-lead stimulus in the PSS did not change. However, in this study, visual apparent motion changed the PSS relative to the sound-lead stimulus, which is closer to physical simultaneity than in the abovementioned studies. This may mean that visual apparent motion contributes to very precise perceptions of temporal simultaneity, which is closer to physical simultaneity, in audiovisual integration.

With respect to temporal resolution, we found that visual apparent motion resulted in greater temporal discrimination. The JND is regarded as an indicator of the temporal window of sensory integration, because it represents the resolution of temporal discrimination between the senses. The JND is known to be in the range of 30–60 ms in audiovisual TOJ tasks using a set of simple stimuli, such as a pair consisting of a single sound and single flash (Morein-Zamir S, et al., 2003; Zampini, et al., 2003; Keetels and Vroomen, 2005). On the other hand, it has been reported that the JND in audiovisual speech is greater (a temporal window of approximately 200 ms) than that in the above-described TOJ tasks in audiovisual integration (Wassenhove. et al., 2007). In particular, previous studies have reported that this temporal resolution changes according to a variety of factors such as spatial or temporal separation of stimuli and predictions regarding the presentation of stimuli (Allik and Kreegipuu, 1998; Jakowśki and Verleger, 2000; Mattes and Ulrich, 1998; Spence, et al., 2001; Schneider and Bavelier, 2003; Spence and Parise, 2010).

For example, the temporal ventriloquism effect shifted the JND toward smaller values than those of the baseline (Morein-Zamir S, et al., 2003, see Table 2-1). On the other hand, in audiovisual speech, the JND of congruent audiovisual speech stimuli (a mean JND of 205 ms) is larger than that of incongruent audiovisual speech such as that under McGurk effect (with a mean JND of 161 ms; see Table 2-1) (Wassenhove. et al., 2007). With respect to spatial separation, Zampini et al. (2003) reported that temporal resolution in audiovisual integration was improved when audiovisual stimuli were presented in different locations rather than in the same location (see Table 2-1) (Zampini, et al., 2003). However, there remained a need to examine whether motion information influences temporal resolution in audiovisual processing. In this study, apparent motion shows greater temporal resolution than that which occurs in the normal condition. Therefore, visual apparent motion is a new factor that increases temporal discrimination.

Table 2-1. Changes in PSS and JND in audiovisual integration in previous studies.

	PSS		JND	
Temporal ventriloquism	baseline	AVVA	baseline	AVVA
	Not changed		62 ms	45 ms
Audiovisual speech	congruent	incongruent	congruent	incongruent
	23 ms	37 ms	205 ms	161 ms
Audiovisual spatial position	same	different	same	different
	60 ms	75 ms	32 ms	22 ms

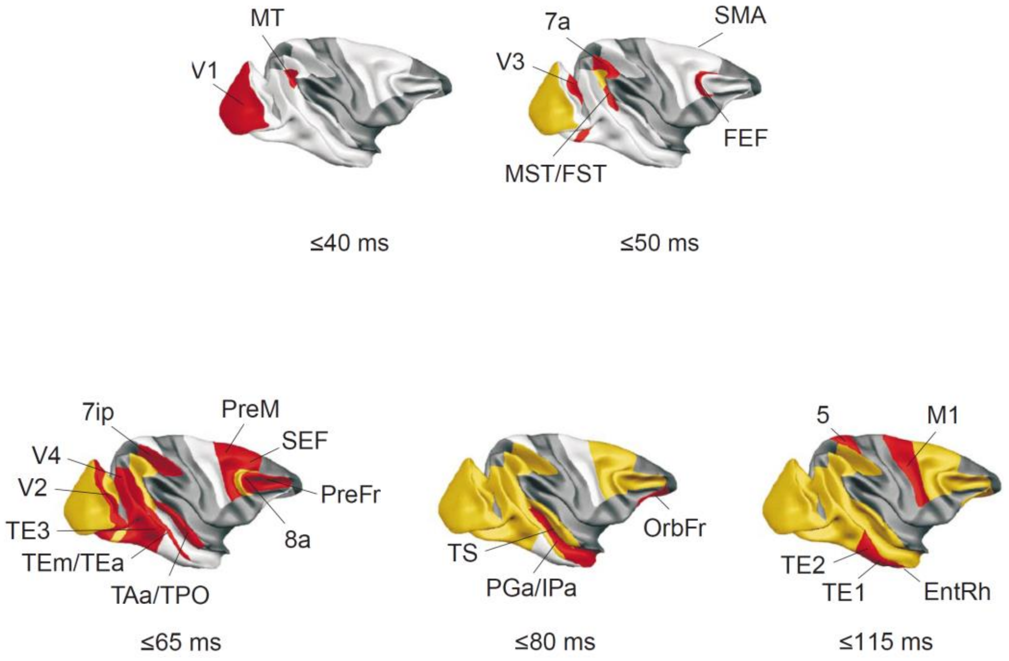
Temporal ventriloquism: "baseline" means that each auditory stimulus was presented at the same time as a flash (i.e., each pair of audiovisual stimuli were presented simultaneously), and "AVVA" indicates that one sound was presented before the first flash and the other sound after the second flash. Audiovisual speech: "congruent" means that auditory speech stimuli were congruent with visual speech stimuli, and "incongruent" indicates that auditory speech stimuli were not congruent with visual speech stimuli, as in McGurk effect speech. Audiovisual spatial position: "same" means the same spatial location, and "different" indicates a different spatial location for audiovisual stimuli.

There is a robust illusory phenomenon called flash-lag effect, which means that motion is perceived faster than a single flash (Nijhawan, 1994; Nijhawan, 2002; Vroomen and Gelder, 2004). When a moving target and a static flash are aligned and when they appeared at the same location, the moving target is perceived as shifted more slightly in the direction of motion relative to the flash. In other words, the flash is processed more slowly than the moving target. However, Brenner and Smeets (2000) pointed out the problem of flash-lag experiment. Brenner and Smeets (2000) claimed that a moment for spatial alignment of moving object and a flash, especially moving target's location, can only be referred after the flash's presentation. Besides, Vroomen and Gelder (2004) examined whether a sound presented with a flash at the same time has an effect on the flash-lag effect. The results showed that when the sound was present with a flash, the flash-lag effect were reduced relative to the silent condition. However, there is a possibility that participants might judge the position of a flash by the single sound as pointed out by Brenner and Smeets (2000). Therefore, although Vroomen and Gelder (2004) confirmed the temporal ventriloquism effect (a single sound presented simultaneously with a flash compensated for the flash-lag effect), it was unclear that whether the single sound attracted the single flash or determined the moment of spatial location of moving object. As mentioned above, flash-lag experiment is that ascertain the relative location of moving object after a flash is perceived. Therefore, there was need to consider dividing motion and a single flash. In the present study, we examined the temporal characteristics of audiovisual integration by dividing the visual apparent motion and a single flash. As a result, the temporal characteristics between visual apparent motion and a single flash were different in audiovisual processing. Visual apparent motion showed that the point of subjective simultaneity (PSS) was shifted toward a sound-lead stimulus, especially became closer to physical simultaneity (i.e., zero) and the just noticeable difference (JND) was reduced compared with a normal TOJ task with a single flash. This indicates that visual apparent motion affects audiovisual simultaneity and improves temporal discrimination in audiovisual processing. Therefore, we quantified the influence of motion information on the temporal integration of multisensory perceptions.

With regard to visual apparent motion and a single sound, it has been reported that the impression of visual apparent motion was changed by a single sound (Getzmann, 2007; Bruns and Getzmann, 2008). Getzmann (2007) have approached the study by a perspective of Gestalt that sensory stimuli are perceived as an overall information. Including no sound condition as control condition, there were four sound conditions; two sounds were presented either before and after the lights as clicks apart or at the same time as simultaneous clicks and intervening between the lights as double clicks or a single click. As a result, the impression of visual apparent motion was improved by a single click that presented temporally in the middle of two lights. On the other hand, clicks apart and simultaneous clicks have resulted in the opposite effect. This finding implicates the possibility that the single sound facilitated the perception of visual apparent motion by attracting or connecting the two lights. Also, Bruns and Getzmann (2008) has demonstrated that visual events are likely to be integrated with the temporally close auditory stimuli during perception of visual apparent motion, especially perception of apparent motion enhanced when auditory stimulation was presented between visual stimuli as a intervening click. Although those cognitive changes may be due to auditory effects, the intrinsic characteristics of apparent motion in time perception are not fully clarified. According to these reports, perception of visual apparent motion could be altered by single and irrelevant sounds, and therefore auditory stimuli were considered to be influential in the temporal structure (Getzmann, 2007; Bruns and Getzmann, 2008). However, it is difficult to clarify the temporal perceptual characteristics of visual apparent motion because they studied only the impression of visual apparent motion by using temporal ventriloquism effect. The temporal perception by spatial shift which is a fundamental property of motion is not investigated. In this study, visual apparent motion had an effect on the audiovisual temporal perception compared with a single flash condition, not affected by a single sound. From the standpoint of motion perception, this finding provides new evidence that there exists a different multisensory process depending on the presence or absence of visual apparent motion.

Our findings revealed that visual apparent motion affects audiovisual simultaneity by shifting simultaneity toward a sound-lead stimulus and physical simultaneity, and reduces the temporal window in audiovisual integration by improving temporal discrimination in audiovisual processing. In physiological aspect, these results may be associated with that motion information has faster and accurate audiovisual processing and that there exists a different multisensory processing depending on the presence or absence of apparent motion. The response of visual apparent motion is typically equivalent to the response of real motion in the physiological mechanisms. Visual apparent motion passes through the dorsal stream associated with visual motion processing (Newsome, et al., 1986; Manning, et al., 1988). In particular, area MT (middle temporal/ V5) plays an important role in motion processing (Born and Bradley, 2005) and especially area MT in the dorsal stream has a function which comprehensively processes the whole motion information of objects by integrating the information of local motion (Pack and Born, 2001). In addition, area MT is greatly implicated in precise spatiotemporal encoding with respect to the timing and interval of visual stimuli (Bair and Koch, 1996; Buracas, et al., 1998; Jantzen, et al., 2005; Bueti, et al., 2008).

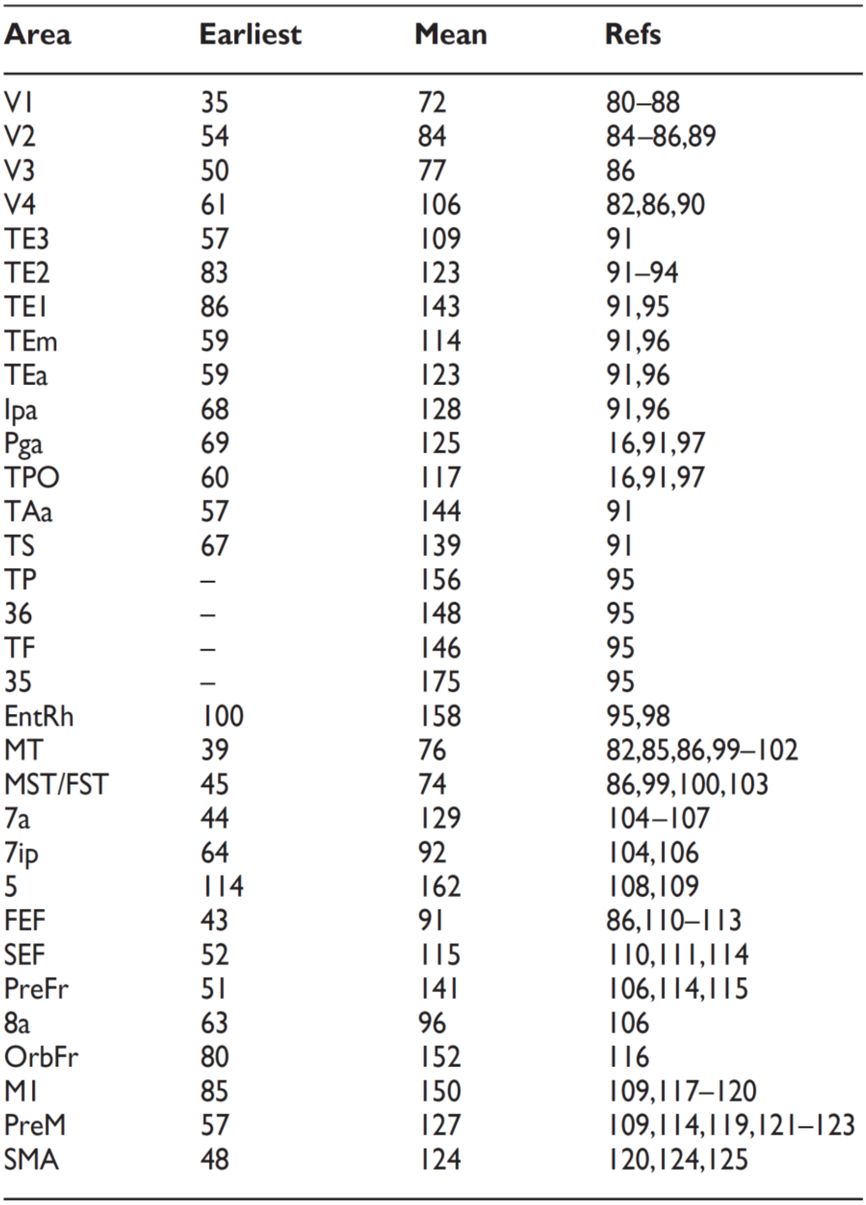
With respect to processing of audiovisual integration, functional imaging studies have found the brain areas of audiovisual integration and, in particular superior temporal sulcus (STS) has been suggested as an audiovisual association area (Stein and Stanford, 2008; Ghazanfar and Schroeder, 2006; Beauchamp, 2005; Olson, et al., 2002). Although area MT has been known as a level in hierarchy of visual processing, several functional imaging studies have found auditory influences on area MT (Ciaramitaro, et al., 2007; Calvert, et al., 1999; Scheef, et al., 2009; Saenz, et al., 2008; Alink, et al., 2008). They suggest that area MT is implicated in multisensory processing. Also, area MT has the fastest response latencies in visual response in macaque cerebral cortex ($\leq 40ms$) and TPO corresponding to the human STS, which has slower response latencies relative to area MT ($\leq 65ms$) (Lamme and Roelfsema, 2000; see Figure 2.4 and Table 2-2). Along with the results of PSS in the present study, we suggest that there exists the difference in the multisensory characteristics between motion information and non-motion information, and especially area MT may contribute to the audiovisual process of motion information.



trends in Neurosciences

Figure 2.4. Earliest visual response latencies of areas in the macaque cerebral cortex (Lamme and Roelfsema, 2000). Figure 2.4 shows activation of areas at specific latencies in MRI images of monkey cortex. Areas that have become active at the given latency after visual stimulation are shown in red in each plot. Regions shown in yellow represent the areas that were activated earlier. White regions are not yet activated. Dark gray regions represent areas for which no information was obtained. Latencies of the layers of the LGN, magnocellular (earliest 28 ms, mean 33 ms) and parvocellular (earliest 31 ms, mean 50 ms) are not shown.

Table 2- 2. Visual response latencies in macaque cerebral cortex (Lamme and Roelfsema, 2000).



Area	Earliest	Mean	Refs
V1	35	72	80–88
V2	54	84	84–86,89
V3	50	77	86
V4	61	106	82,86,90
TE3	57	109	91
TE2	83	123	91–94
TE1	86	143	91,95
TEm	59	114	91,96
TEa	59	123	91,96
Ipa	68	128	91,96
Pga	69	125	16,91,97
TPO	60	117	16,91,97
TAa	57	144	91
TS	67	139	91
TP	–	156	95
36	–	148	95
TF	–	146	95
35	–	175	95
EntRh	100	158	95,98
MT	39	76	82,85,86,99–102
MST/FST	45	74	86,99,100,103
7a	44	129	104–107
7ip	64	92	104,106
5	114	162	108,109
FEF	43	91	86,110–113
SEF	52	115	110,111,114
PreFr	51	141	106,114,115
8a	63	96	106
OrbFr	80	152	116
M1	85	150	109,117–120
PreM	57	127	109,114,119,121–123
SMA	48	124	120,124,125

In summary, we investigated whether visual apparent motion affects audiovisual simultaneity perception. Participants performed temporal order judgment (TOJ) tasks between apparent motion condition and normal condition. As a result, our findings reveal that visual apparent motion affects audiovisual simultaneity by shifting simultaneity toward physical simultaneity, and reduces the temporal window in audiovisual integration. This suggests that visual motion information contributes to accurate perceptions of temporal events in the physical world.

Chapter 3

The Mechanism of Effect of Motion

Chapter 2 was focused on whether visual apparent motion affects audiovisual simultaneity perception. Participants performed temporal order judgment (TOJ) tasks between the apparent motion condition and normal condition. As a result, the findings reveal that visual apparent motion affects audiovisual simultaneity by shifting simultaneity toward physical simultaneity, and reduces the temporal window in audiovisual integration. However, it is not clear that what mechanisms contribute to the finding that visual apparent motion affects audiovisual simultaneity perception. Therefore, there is a need to confirm the possible mechanisms. In the results of Chapter 2, three possible mechanisms can be considered. First, it remained unclear whether the results were influenced by the different amount of visual stimulation between the apparent motion condition with two flashes and the normal condition with a single flash. Motion cannot be perceived with a single flash because two flashes are a fundamental unit for perceiving motion. Therefore, there exist the difference of the amount of visual stimulation between motion and a single flash. Therefore, there is a need to examine the influence of the amount of visual stimulation. Second, because apparent motion was produced by a constant interval between two flashes, the results may be accounted for by specific prediction. Therefore, it was necessary to conduct a supplementary experiment in which the interval was changed randomly. Third, binding property is essential for motion perception in human perceptual system. Apparent motion is a most suitable phenomenon to show motion binding property in human perceptual system. Therefore, there is a need to examine the influence of motion binding property. In Chapter 3, we consisted of three sections for examining the possible mechanisms. In Section 1 of Chapter 3, we examined the influence of the amount of visual stimulation by eliminating the influence of the amount of flash stimulation. In Section 2 of Chapter 3, we examined the influence of prediction by a constant interval between two flashes using randomizing the intervals between the two flashes. In Section 3 of Chapter 3, we examined the influence of motion binding property.

3.1

The Influence of Amount of Stimulation

3.1.1 Introduction

Chapter 2 revealed the influence of visual apparent motion on an audiovisual TOJ task. Participants conducted audiovisual TOJ tasks in the apparent motion condition with two flashes, and in the normal condition with a single flash, which is the conventional condition of a TOJ task. However, it was insufficient only to compare the apparent motion condition with the normal condition because the two conditions in Chapter 2 differ in the amount of visual stimulation (see Figure 3.1). In previous studies, larger amount of visual stimulation increase visual energy and prime the visual processing system (Spence, et al., 2001; Posner and Nissen, 1976). Therefore, there is a need to examine the effect of visual apparent motion compared with visual stimulation of identical amount. Section 1 consisted of two kinds of TOJ tasks in the apparent motion condition and in the successive condition. In previous studies, when the temporal interval between two flashes was long, visual apparent motion was not detected; the two flashes were perceived as successive events with no movement (Getzmann, 2007; Harrar, et al., 2008; Strybel, et al., 1990). For example, many researchers have reported that two visual stimuli are perceived as a continuous motion when the interstimulus onset interval (ISOI) of the visual stimuli is within a range of 50 to 150 ms and visual angle is one or more degree. Conversely, the visual stimuli are perceived as successive beyond an ISOI of 300 ms (Dawson, 1991; Briggs and Perrott, 1972; Getzmann, 2007; Harrar, et al., 2008; Strybel, et al., 1990). Therefore, participants preformed TOJ tasks during apparent motion perception and successive perception without motion.

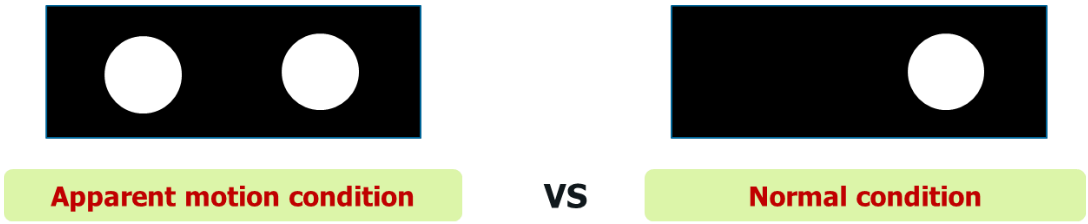


Figure 3.1. Comparison in stimulus structures between the apparent motion condition and normal condition in Chapter 2. The amount of visual stimulation differed between the apparent motion condition with two flashes and the normal condition with a single flash in Chapter 2. Motion cannot be perceived with a single flash because two flashes are a fundamental unit for perceiving motion.

3.1.2 Method

3.1.2.1 Task Design

We conducted audiovisual TOJ tasks in which the spatial location and duration of stimuli were identical. Participants performed the TOJ task in the apparent motion condition in which two flashes were presented with a stimulus onset asynchrony (SOA) of 137 ms (Harrar, et al., 2008), which is the same spatiotemporal interval as in the experiment of Chapter 2, and in the successive condition with SOAs of 300 and 500 ms between the two flashes, which are perceived as successive stimuli without motion (Dawson, 1991; Briggs and Perrott, 1972; Getzmann, 2007; Harrar, et al., 2008; Strybel, et al., 1990).

3.1.2.2 Participants

Twelve participants (10 males and two females, with a mean age of 23.9 years) took part in Section 1. Participants were paid to take part in the experiments, and written informed consent was obtained. These experiments were approved by the ethics committee of the Tokyo Institute of Technology.

3.1.2.3 Apparatus and Stimuli

In Section 1, the apparatus and stimuli was the same as in Chapter 2.

3.1.2.4 Procedure

In Section 1, the procedure was the same as in Chapter 2 with the following exceptions. Section 1 consisted of two kinds of TOJ tasks in the apparent motion condition (Figure 3.2 (A)) and the successive condition (Figure 3.2 (B)). The apparent motion condition was equivalent to that of Chapter 2, with an SOA of 137 ms between the two flashes. The successive condition consisted of SOAs of 300 or 500 ms between the two flashes. The timing of the auditory stimulus relative to the second flash was the same as in Chapter 2. The participants were instructed to judge the order of the second visual stimulus and the brief sound. Section 1 consisted of 405 trials (3 visual conditions × 9 audiovisual SOAs × 15 repetitions) in counterbalanced order and the experiment was divided into nine blocks of 45 trials (9 audiovisual SOAs × 5 repetitions). Including practice for each task, each experiment took approximately one and a half hours.

Before starting each experiment, we examined whether the participants perceived motion between two flashes. In Section 1, we showed the participants two flashes with an SOA of 137 ms 10 times, and participants were asked to evaluate whether their impression of the stimuli was of “continuous motion” or “successive stimuli.” In Section 1, we showed them two flashes, with SOAs of 137 ms, 300 ms and 500 ms, 10 times in counterbalanced order. Then, the participants were asked to evaluate whether their impression of the stimuli was of “continuous motion” or “successive stimuli.” Only the participants who perceived “continuous motion” with an SOA of 137 ms and “successive stimuli” with SOAs of 300 ms and 500 ms for all stimuli continued to participate in the experiment. We confirmed that motion was perceived during the TOJ task after each experimental session was completed.

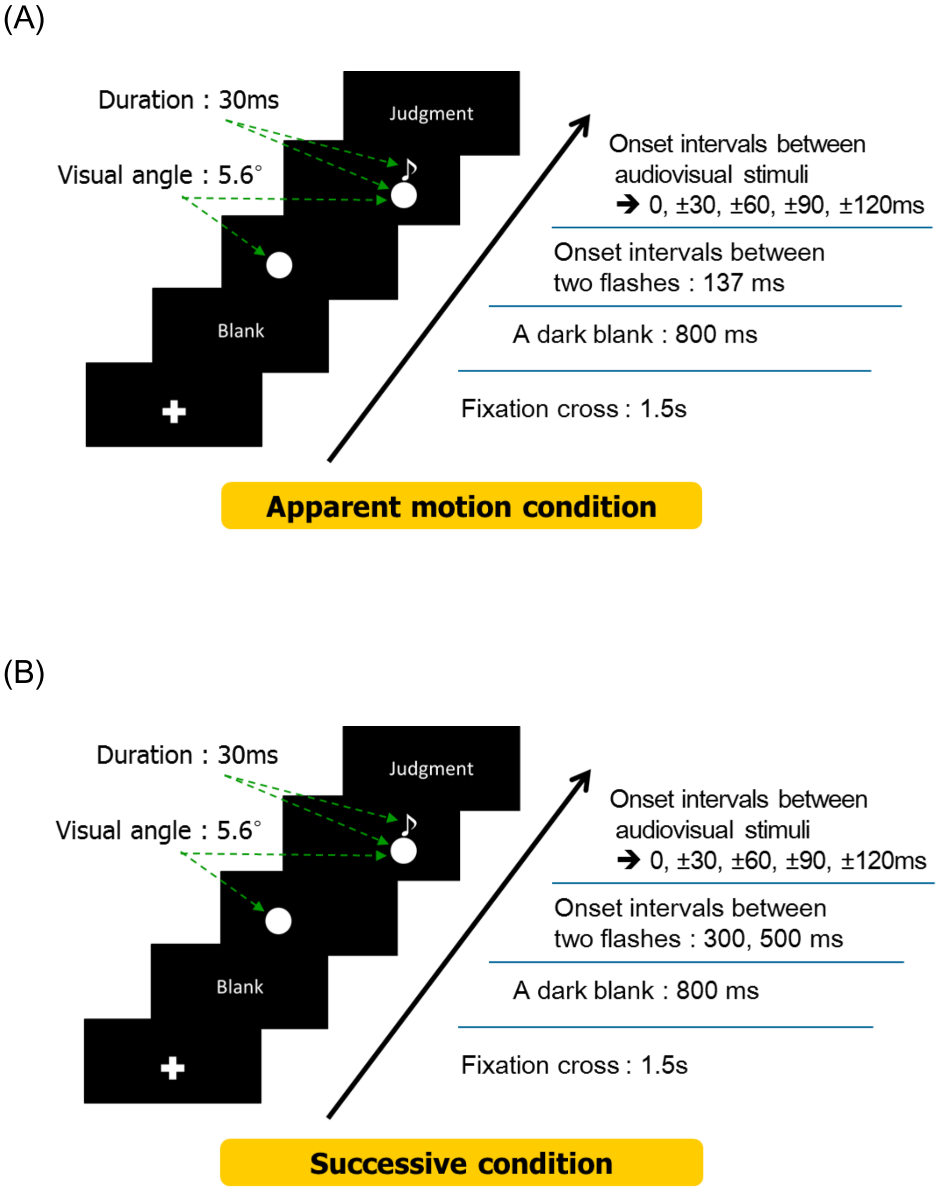


Figure 3.2. Schematic illustration of Section 2. (A) and (B) show the procedure for Section 1. After the presentation of a fixation cross (1.5 s) and a blank screen (800 ms), two visual stimuli were presented in the apparent motion condition (A) and two visual stimuli with SOAs of 300 and 500 ms in the successive condition (B). Sounds were presented either before or after the visual stimuli at SOAs ranging from –120 to 120 ms at 30 ms intervals in random order.

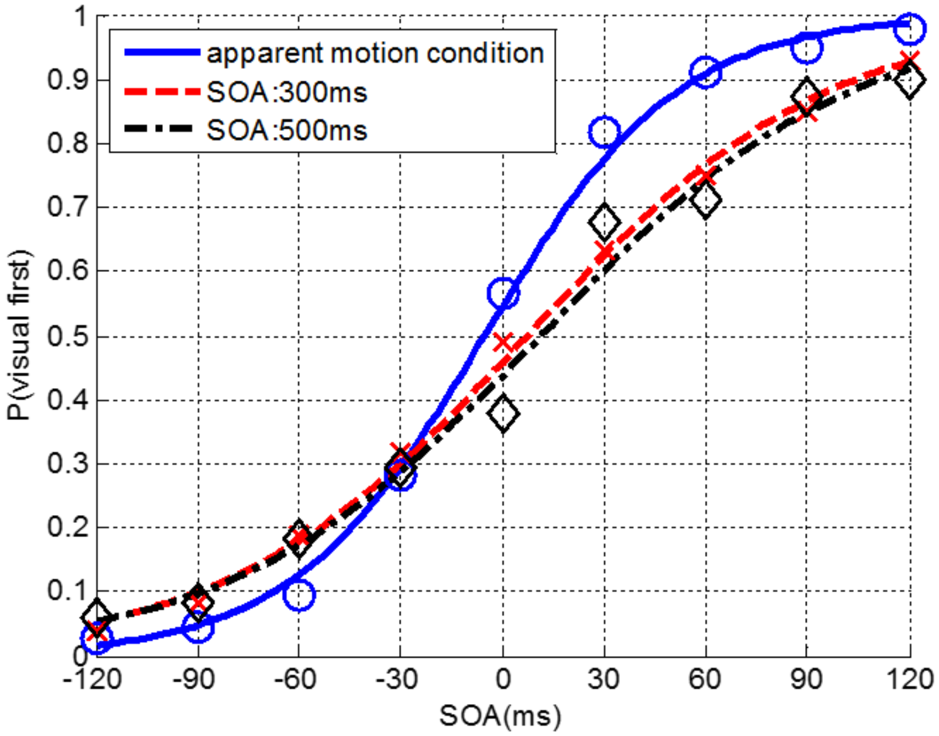
3.1.2.5 Data Analysis

In Section 1, the Data analysis was the same as in Chapter 2.

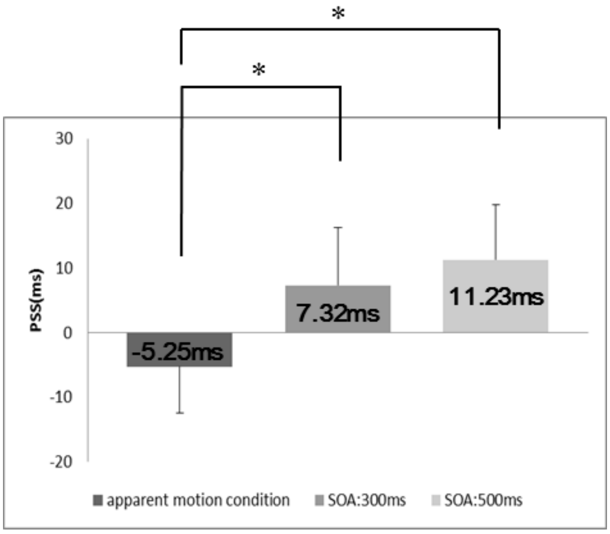
3.1.3 Results

In Section 1, the participants performed both kinds of TOJ tasks in the apparent motion and successive conditions. Individual PSSs and JNDs for the two conditions were computed as in Chapter 2. Figure 3.3 (A) illustrates the results of Section 1 with psychometric curves fitted to the distribution of the mean TOJ data of all participants in the apparent motion and successive conditions. Figures 3.3 (B) and (C) show the results of PSS and JND in Section 1. The PSS shifted toward a sound-lead stimulus and became closer to physical simultaneity (i.e., zero) in the apparent motion condition, and the JND in the apparent motion condition was smaller than that in the successive condition. In particular, the PSS and JND in the apparent motion condition were almost identical to those obtained in the apparent motion condition in Chapter 2, and similar results were obtained in the successive condition in Section 1 and in the normal condition in Chapter 2. A repeated-measures analysis of variance (ANOVA) of the PSSs showed a significant main effect for the temporal interval, $F(2, 23) = 15.83$, $p < 0.001$ (see Table A-3). Multiple comparisons with Holm correction showed significant differences between the apparent motion condition and the successive condition (apparent motion and an SOA of 300 ms: $p = 0.042$, apparent motion and an SOA of 500 ms: $p = 0.012$, see Table A-3). Moreover, a repeated-measures ANOVA of the JNDs revealed a significant main effect of the temporal interval, $F(2, 23) = 25.03$, $p < 0.001$ (see Table A-4). In addition, multiple comparisons with Holm correction confirmed that there was a significant difference between the apparent motion and successive conditions (apparent motion and an SOA of 300 ms: $p = 0.002$, apparent motion and an SOA of 500 ms: $p < 0.001$, see Table A-4).

(A)


apparent motion condition
SOA:300ms
SOA:500ms
P(visual first)
SOA(ms)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
-120
-90
-60
-30
0
30
60
90
120

(B)


*
*
PSS(ms)
30
20
10
0
-10
-20
-5.25ms
7.32ms
11.23ms
apparent motion condition
SOA:300ms
SOA:500ms

(C)

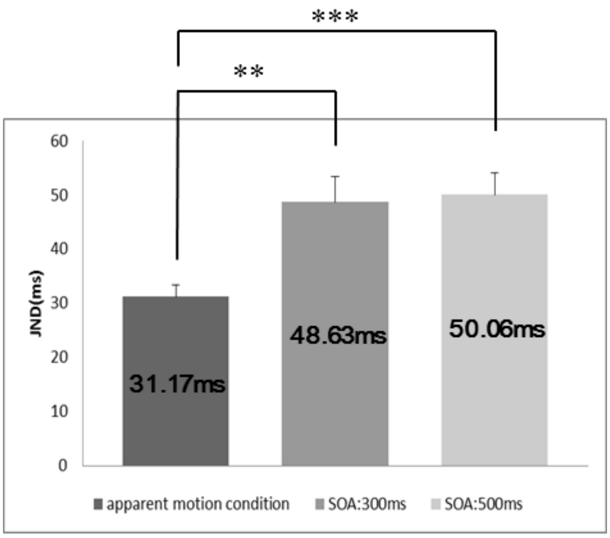


Figure 3.3. Results of Section 1. (A) shows psychometric curves fitted to the distribution of mean TOJ data for all participants in the apparent motion and successive conditions. The negative values of the SOAs indicate that an auditory stimulus preceded a visual stimulus. (B) shows the mean PSSs in the apparent motion and successive conditions. Error bars represent the standard error of the means. (C) shows the mean JNDs in the apparent motion and successive conditions (SOA: 300 ms and 500 ms). Error bars represent the standard error of the means. The *p* values were adjusted for multiple comparisons using Holm’s correction, *: *p* < .05, **: *p* < .01, ***: *p* < .001.

3.1.4 Discussion

In Section 1, the results obtained in the apparent motion condition differ from those obtained in the successive condition, which included the same amount of visual stimulation in the apparent motion condition. Similar results were obtained in the apparent motion conditions between Chapter 2 and Section 1 in Chapter 3. Also, the results obtained in the successive condition were similar to those obtained in the normal condition in Chapter 2. This means that the PSSs and JNDs in audiovisual temporal perception differed according to whether visual apparent motion was present or absent. In Chapter 2, it remained unclear whether the results were influenced by differences in the amount of visual stimulation between the apparent motion condition with two flashes and the normal condition with a single flash. There was a possibility that the amount of visual apparent motion influenced audiovisual simultaneity perception because visual energy increases by larger amount in the apparent motion condition (Posner and Nissen, 1976). In addition, because double visual stimuli in the apparent motion condition would prime the visual processing system, this may change the relative perception of audiovisual simultaneity compared with the use of a single visual stimulus (Spence, et al., 2001; Posner and Nissen, 1976). It is particularly notable that Spence et al. (2001) reported that the temporal processing of modalities was affected not only by the prediction of modality expectancies but also by the quantity of modality-induced stimuli. In Section 1, therefore, the amount of visual stimulation in the two conditions was equalized, and the participants then conducted TOJ tasks in the apparent motion condition and in the successive condition. However, the amount of visual stimulation made no difference in the apparent motion condition.

Visual apparent motion is systematically affected by the temporal interval between two stimuli. When the temporal interval is too short, the stimuli are perceived as simultaneous, whereas when the temporal interval exceeds a certain interval, the two stimuli are perceived as successive. Therefore, if the temporal interval between two stimuli is too short or too long, motion is not perceived. In this section, we examined the simultaneity perception depending on the presence or absence of apparent motion. As a result, the PSS shifted toward a sound-lead stimulus and toward physical simultaneity in the apparent motion condition, and the JND in the apparent motion condition was smaller than that in the successive condition. In particular, the PSS and JND in the apparent motion condition were almost identical to those obtained in the apparent motion condition in Chapter 2, and similar results were obtained in the successive condition in Section 1 and in the normal condition in Chapter 2. This findings may mean that the simultaneity perception of motion information differ from that of non-motion information.

3.2

The Influence of Prediction

3.2.1 Introduction

In section 1, it became clear that the result obtained under the apparent motion condition was unaffected by the amount of flash stimulation. However, in Chapter 2, the results may be accounted for by specific prediction, because apparent motion was produced by a constant interval between two flashes. Therefore, there may remain an influence not only of apparent motion, but also of specific prediction as a higher-order brain function. With respect to visual prediction and attention, it should be noted that when participants know the specific time at which targets appear, specific attention can be allocated (Coull and Nobre, 1998). Furthermore, predictable and anticipated information improves temporal resolution and temporal sensitivity (Petrini, et al., 2009). Therefore, it was necessary to eliminate the influence of prediction by randomizing the intervals between the two visual stimuli. In this section, we removed the influence of prediction by randomizing the intervals between the two visual stimuli.

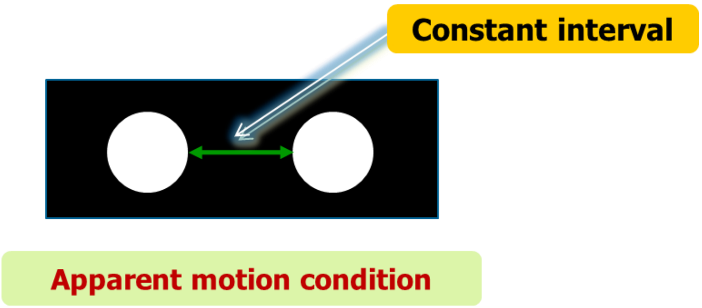


Figure 3.4. The influence of prediction during apparent motion perception. Apparent motion was produced by a constant interval between two flashes in Chapter 2.

3.2.2 Method

3.2.2.1 Task Design

We conducted audiovisual TOJ tasks in which the spatial location and duration of stimuli were identical. In Section 2, we wished to eliminate the effect of prediction because of constant intervals. For that reason, we set up two kinds of TOJ tasks in the apparent motion condition with an SOA of 137 ms, which is the same spatiotemporal interval as in Chapter 2, and in the successive condition with SOAs of 300 and 500 ms between the two flashes, which are perceived as successive stimuli without motion. We presented the two visual stimuli with SOAs of 137, 300, or 500 ms in random order to eliminate the effect of prediction.

3.2.2.2 Participants

Twelve participants (11 males and one female, with a mean age of 23.5 years) took part in Section 2. Participants were paid to take part in the experiments, and written informed consent was obtained. These experiments were approved by the ethics committee of the Tokyo Institute of Technology.

3.2.2.3 Apparatus and Stimuli

In Section 2, the apparatus and stimuli was the same as in Chapter 2.

3.2.2.4 Procedure

In Section 2, the procedure was the same as in Chapter 2, with the following exceptions. In Section 2, the participants conducted the TOJ tasks with SOAs of 137 ms, 300 ms, and 500 ms presented in random order between the visual stimuli to confirm the influence of prediction by a constant interval (see Figure 3.5). The timing of the auditory stimulus presented with the second flash was the same as that in Chapter 2. The participants were instructed to judge the order of the second visual stimulus and the brief sound. Section 2 consisted of 432 trials (3 visual conditions

× 9 audiovisual SOAs × 16 repetitions) in counterbalanced order and the experiment was divided into eight blocks of 54 trials (3 visual conditions × 9 audiovisual SOAs × 2 repetitions). Including practice for each task, each experiment took approximately one and a half hours.

Before starting each experiment, we examined whether the participants perceived motion between two flashes. In Section 2, we showed the participants two flashes with an SOA of 137 ms 10 times, and participants were asked to evaluate whether their impression of the stimuli was of “continuous motion” or “successive stimuli.” In Section 2, we showed them two flashes, with SOAs of 137 ms, 300 ms and 500 ms, 10 times in counterbalanced order. Then, the participants were asked to evaluate whether their impression of the stimuli was of “continuous motion” or “successive stimuli.” Only the participants who perceived “continuous motion” with an SOA of 137 ms and “successive stimuli” with SOAs of 300 ms and 500 ms for all stimuli continued to participate in the experiment. We confirmed that motion was perceived during the TOJ task after each experimental session was completed.

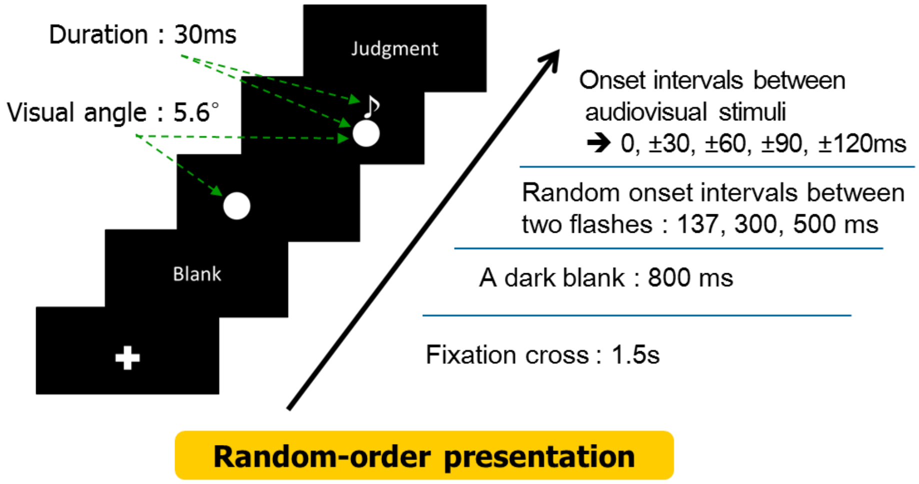


Figure 3.5. Schematic illustration of Section 2. Figure 3.5 shows the procedure for this experiment. After the presentation of a fixation cross (1.5 s) and a blank screen (800 ms), two visual stimuli were presented with SOAs of 137, 300 and 500 ms in random order (random order condition). In all the experiments, sounds were presented either before or after the visual stimuli at SOAs ranging from –120 to 120 ms at 30 ms intervals in random order. Sounds were presented either before or after the visual stimuli at SOAs ranging from –120 to 120 ms at 30 ms intervals in random order.

3.2.2.5 Data Analysis

In Section 2, the Data analysis was the same as in Chapter 2.

3.2.3 Results

In Section 2, the participants performed the TOJ task with an interval between two visual stimuli that was varied in random order, and only the results obtained in the apparent motion condition (with an SOA of 137 ms) were extracted. All the participants perceived continuous motion, and the PSSs and JNDs were computed as in Chapter 2. Figure 3.6 (A) shows the results of Section 2, and Figures 3.6 (B) and (C) show the results for the PSSs and JNDs in Section 2. The nonrandom-order presentation reflects the results of the apparent motion condition in Chapter 2. The values of the PSS and JND in the apparent motion condition in Section 2 were almost the same as those obtained in the apparent motion condition in Chapter 2. An unpaired t-test of PSSs and JNDs for the TOJ tasks in the apparent motion condition indicated no significant difference between Chapter 2 and Section 2 in Chapter 2 ($t(26) = -0.11$, $p = 0.92$, $t(26) = -0.12$, $p = 0.91$, see Tables A-5 and A-6).

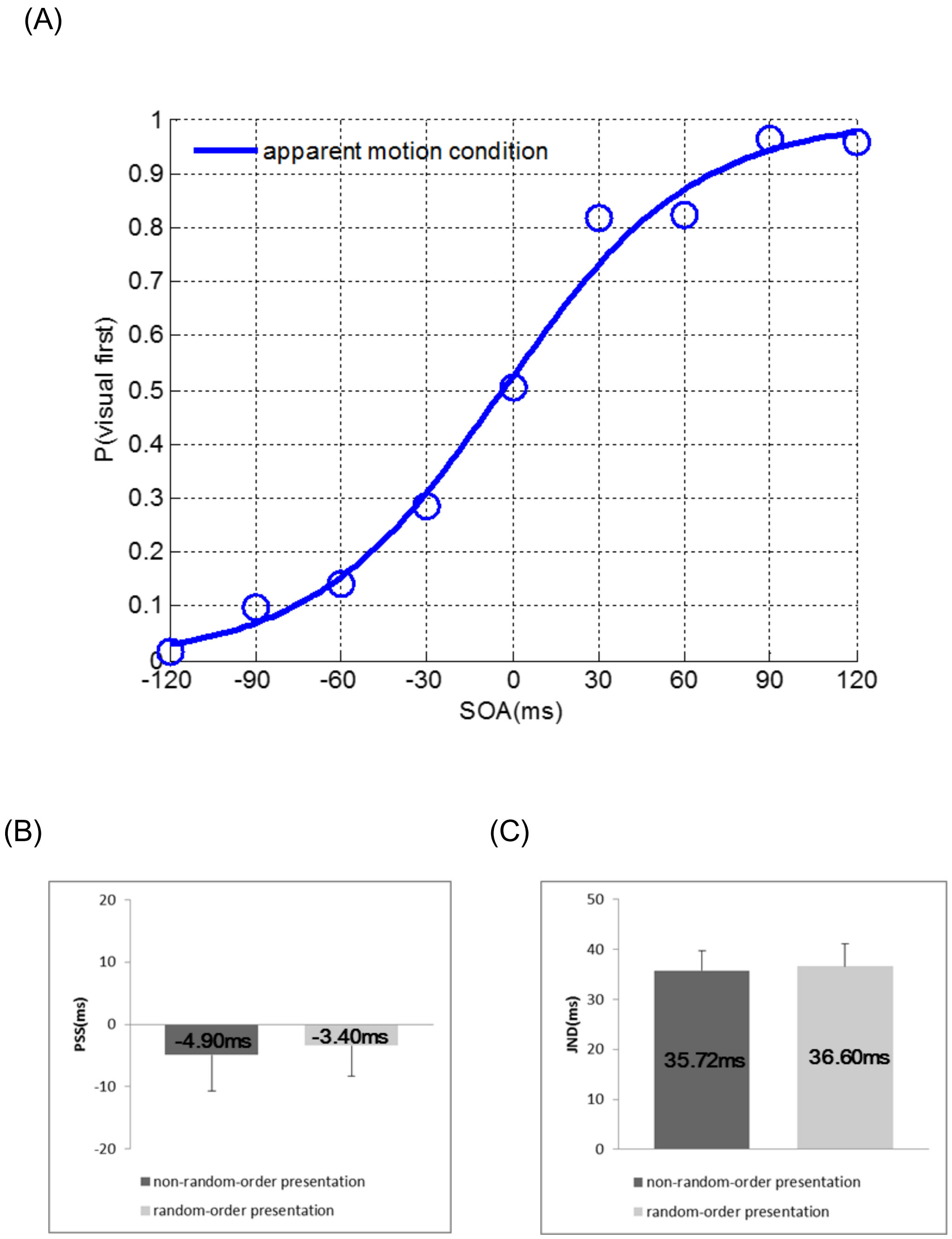


Figure 3.6. Results of Section 2. (A) shows psychometric curves fitted to the distribution of the mean TOJ value for all participants in the random order condition. The negative values of the SOAs indicate that an auditory stimulus preceded a visual stimulus. (B) shows the mean PSSs from presentation at nonrandom intervals (the apparent motion condition in Chapter 2) and random intervals (the apparent motion condition in Section 2 in Chapter 3). Error bars represent the standard error

of the means. (C) shows the mean JNDs of the apparent motion and the random order conditions. Error bars represent the standard error of the means, *: p < .05, **: p < .01, unpaired t-test.

3.2.4 Discussion

The results of Section 2, which eliminated the effect of prediction, were not different from those obtained in the apparent motion condition in Chapter 2. This means that apparent motion was equivalently processed regardless of prediction. With respect to visual prediction and attention, it should be noted that when participants know the time at which targets appear, specific attention can be allocated (Coull and Nobre, 1998). Furthermore, it is known that predictable and anticipated information improves temporal resolution and temporal sensitivity (Petrini, et al., 2009). The attention modulates neural activity (O'Craven, et al., 1997; Treue and Martinez Trujillo, 1999), and a faster time course is allocated for visual attention (Busse, et al., 2008). However, the result of unpredictable apparent motion did not differ from that of predictable apparent motion. Therefore, prediction and intention as top-down factors have no effect on the results of Chapter 2.

Many researchers have reported that prediction influences the temporal processing of modalities because temporal processing of expected modalities is faster than that of unexpected modalities (Posner, et al., 1980; Duncan, 1980; Spence and Driver, 1997; Klein, 1977). However, Spence et al. reported that reaction times for a sensory stimulus followed by another sensory stimulus of the same type were faster than when a cross-modal stimulus was expected (Spence, et al., 2001). Therefore, Spence et al. suggested that stimulus-driven and expectancy-driven effects must be distinguished in studies of the temporal processing of sensory modalities (Spence, et al., 2001). The results of Chapter 2 and Section 2 in Chapter 3 show no influence on prediction in audiovisual temporal processing during apparent motion perception. Therefore, temporal processing during apparent motion perception may result from stimulus-driven effects rather than expectancy-driven effects. On the other hand, in Section 2, although a sensory stimulus was followed by another of the same type of sensory stimulus, audiovisual temporal processing during apparent motion perception differed from that during nonapparent motion perception. This suggests the possibility that despite the stimulus-driven effects with the same type of sensory stimuli, the difference between motion and nonmotion perceptions influenced temporal order perception and the window of temporal integration in audiovisual processing.

Specifically, the pathway of motion and non-motion processing may differ in audiovisual temporal perception, and there may exist some activation for determining the pathway of motion and non-motion processing by bottom-up signals. Many researchers have claimed that audio-visual stimulation was integrated at an early processing stage (Fendrich and Corballis, 2001; Bruns and Getzmann, 2008). Fendrich and Corballis (2001) suggested that the sensory capture phenomenon may be connected with low-level inter-sensory linking processes and it seems that auditory driving or auditory dominance depends on such low level sensory linking processes. Also, Bruns and Getzmann (2008) also reported that their findings are consistent with a low-level audiovisual integration between visual apparent motion and a single sound. Moreover, in recent years, with the growing interest in multisensory properties of motion information, some researchers have raised a possibility that the area MT, which is earlier cortical area and plays an important role in visual motion processing, is engaged in the audiovisual processing (Olson, et al., 2002; Kafaligonul and Stoner, 2010; Calvert, et al., 1999; Born and Bradley, 2005). In this study, the same properties showed between predictable motion (Chapter 2) and unpredictable motion (Section 2 in Chapter 3) on audiovisual temporal processing. Therefore, we suggest that the pathway for motion or non-motion processing is determined by the bottom-up signals regardless of prediction as top-down factors. This fact may lead us to the mechanism that the pathway of motion or non-motion processing on audiovisual temporal perception depends on the bottom-up signals.

3.3

The Influence of the Binding Property

3.3.1 Introduction

In Section 1, it became clear that the result obtained under the apparent motion condition was unaffected by the amount of flash stimulation. Besides, in Section 2, apparent motion was equivalently processed regardless of prediction and therefore prediction as top-down factors have no effect on the results of Chapter 2. Therefore, the amount of flash stimulation and prediction have no effect on the results of the apparent motion condition. What mechanisms contribute to the finding that apparent motion affects temporal perception? One possibility is the peculiar motion perception mechanism in humans (Watson, et al., 1986; Watson, et al., 1985). When two discrete stimuli are presented at appropriate spatiotemporal intervals, we can perceive continuous motion. In other words, two discrete stimuli separated by intervals smaller than 50 ms are perceived as simultaneous stimuli or greater than 300 ms are perceived as successive stimuli, whereas if the interval is within a range of 50 to 150 ms, the stimuli are perceived as one moving object (Getzmann, 2007; Harrar, et al., 2008; Strybel, et al., 1990; see Figure 3.7). This phenomenon shows motion binding property as the mechanism of visual motion perception. Motion binding property indicates that two visual stimuli are perceived as a single bounded and moving object with spatiotemporal continuity at only certain intervals (Dawson, 1991; Treisman, 1996; Kahneman, et al., 1992). Apparent motion is a most suitable phenomenon to show motion binding property in human perceptual system. Our findings raise the possibility that binding property in visual motion perception influences audiovisual simultaneity perception. Therefore, there is a need to examine the influence of motion binding property.

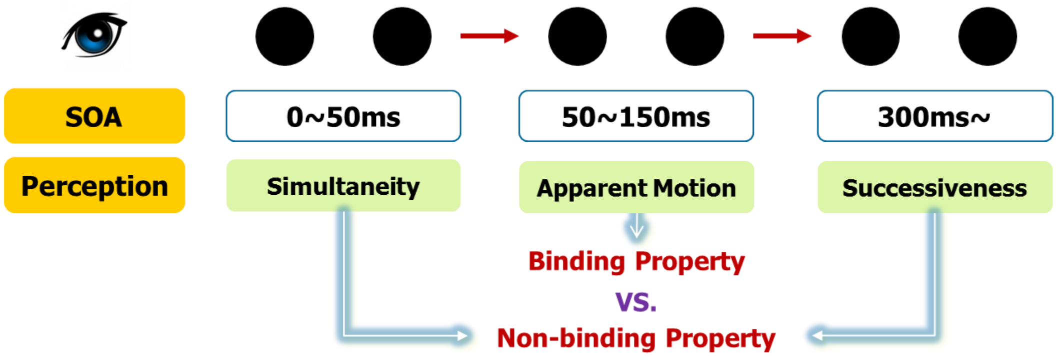


Figure 3.7. Motion binding property. Two discrete stimuli separated by intervals are perceived as simultaneousness, successiveness or motion (Getzmann, 2007; Harrar, et al., 2008; Strybel, et al., 1990). In particular, when two discrete stimuli are presented at appropriate spatiotemporal intervals, we can perceive continuous motion. This phenomenon shows motion binding property that visual stimuli are perceived as a single bounded moving object with spatiotemporal continuity (Dawson, 1991; Treisman, 1996; Kahneman, et al., 1992).

3.3.2 Method

3.3.2.1 Task Design

We conducted audiovisual TOJ tasks in three experiments in which the spatial location and duration of stimuli were identical. In Experiment 1, we set up two kinds of TOJ tasks in the motion binding condition with an SOA of 137 ms, which is the same spatiotemporal interval as in Chapter 2, and in the simultaneous condition with SOA of 0 ms between the two flashes, which are perceived as simultaneous stimuli without motion. In Experiment 2, we set up two kinds of TOJ tasks in the motion binding condition with an SOA of 137 ms and in the successive condition with SOAs of 300 and 500 ms between the two flashes, which are perceived as successive stimuli without motion. In Experiment 3, we set up two kinds of TOJ tasks in the motion binding condition with an SOA of 137 ms, which is the same spatiotemporal interval as in Chapter 2, and in the non-motion binding condition with SOAs of 40 and 300 ms between the two flashes, which are perceived as discrete stimuli without motion binding. The SOA of 40 ms shows the temporal interval with no motion perception and one of the closest SOA to the greatest lower bound the temporal interval for the apparent motion perception. The SOA of 300 ms represents the temporal interval without motion and one of the closest SOA to the least upper bound of the temporal interval for the apparent perception between the two flashes.

3.3.2.2 Participants

Twelve participants (11 males and one female, with a mean age of 25.0 years) took part in Experiment 1. Twelve participants (10 males and two females, with a mean age of 23.9 years) took part in Experiment 2. Twelve participants (11 males and one female, with a mean age of 24.9 years) took part in Experiment 3. Participants were paid to take part in the experiments, and written informed consent was obtained. These experiments were approved by the ethics committee of the Tokyo Institute of Technology.

3.3.2.3 Apparatus and Stimuli

In Chapter 3.3, the apparatus and stimuli was the same as in Chapter 2.

3.3.2.4 Procedure

In Experiment 1, the procedure was the same as in Chapter 2 with the following exceptions. Experiment 1 consisted of two kinds of TOJ tasks in the motion binding condition (Figure 3.8 (A)) and the successive condition (Figure 3.8 (B)). The motion binding condition was equivalent to that of Chapter 2, with an SOA of 137 ms between the two flashes. The simultaneous condition consisted of SOA of 0 ms between the two flashes. The timing of the auditory stimulus relative to the second flash was the same as in Chapter 2. The participants were instructed to judge the order of the second visual stimulus and the brief sound in the motion binding condition. In the simultaneous condition, the participants judged the order of the right flash and the brief sound. Experiment 1 consisted of 270 trials (2 visual conditions × 9 audiovisual SOAs × 15 repetitions) in counterbalanced order and the experiment was divided into 10 blocks of 27 trials (9 audiovisual SOAs × 3 repetitions). Including practice for each task, each experiment took approximately one and a half hours.

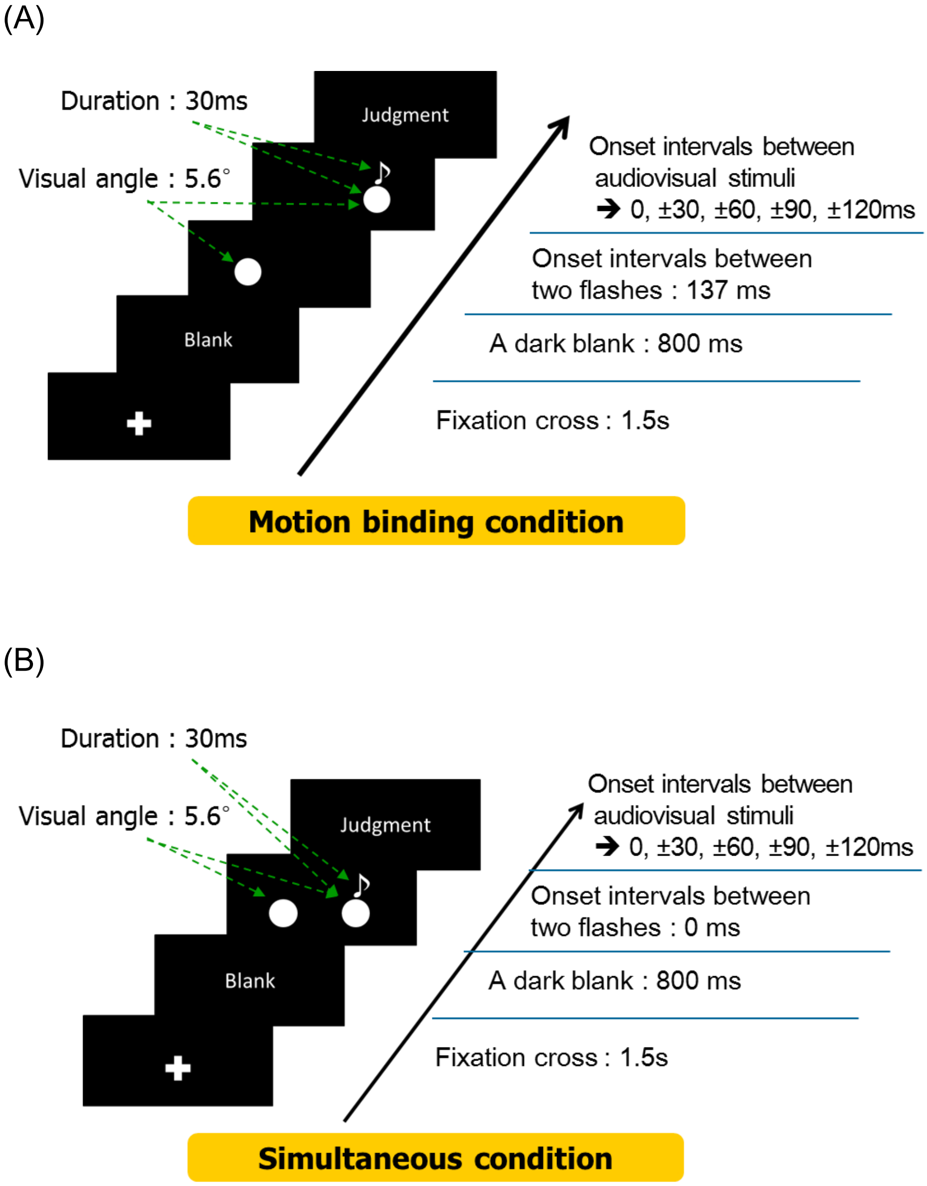


Figure 3.8. Schematic illustration of Experiment 1. (A) and (B) show the procedure for Experiment 1. After the presentation of a fixation cross (1.5 s) and a blank screen (800 ms), two visual stimuli were presented in the motion binding condition (A) and two visual stimuli with SOA of 0 ms in the simultaneous condition (B). Sounds were presented either before or after the visual stimuli at SOAs ranging from –120 to 120 ms at 30 ms intervals in random order.

In Experiment 2, the procedure was the same as in Chapter 2 with the following exceptions. Experiment 2 consisted of two kinds of TOJ tasks in the motion binding condition (Figure 3.9 (A)) and the successive condition (Figure 3.9 (B)). The motion binding condition was equivalent to that of Chapter 2, with an SOA of 137 ms between the two flashes. The successive condition consisted of SOAs of 300 or 500 ms between the two flashes. The timing of the auditory stimulus relative to the second flash was the same as in Chapter 2. The participants were instructed to judge the order of the second visual stimulus and the brief sound. Experiment 2 consisted of 405 trials (3 visual conditions × 9 audiovisual SOAs × 15 repetitions) in counterbalanced order and the experiment was divided into nine blocks of 45 trials (9 audiovisual SOAs × 5 repetitions). Including practice for each task, each experiment took approximately one and a half hours.

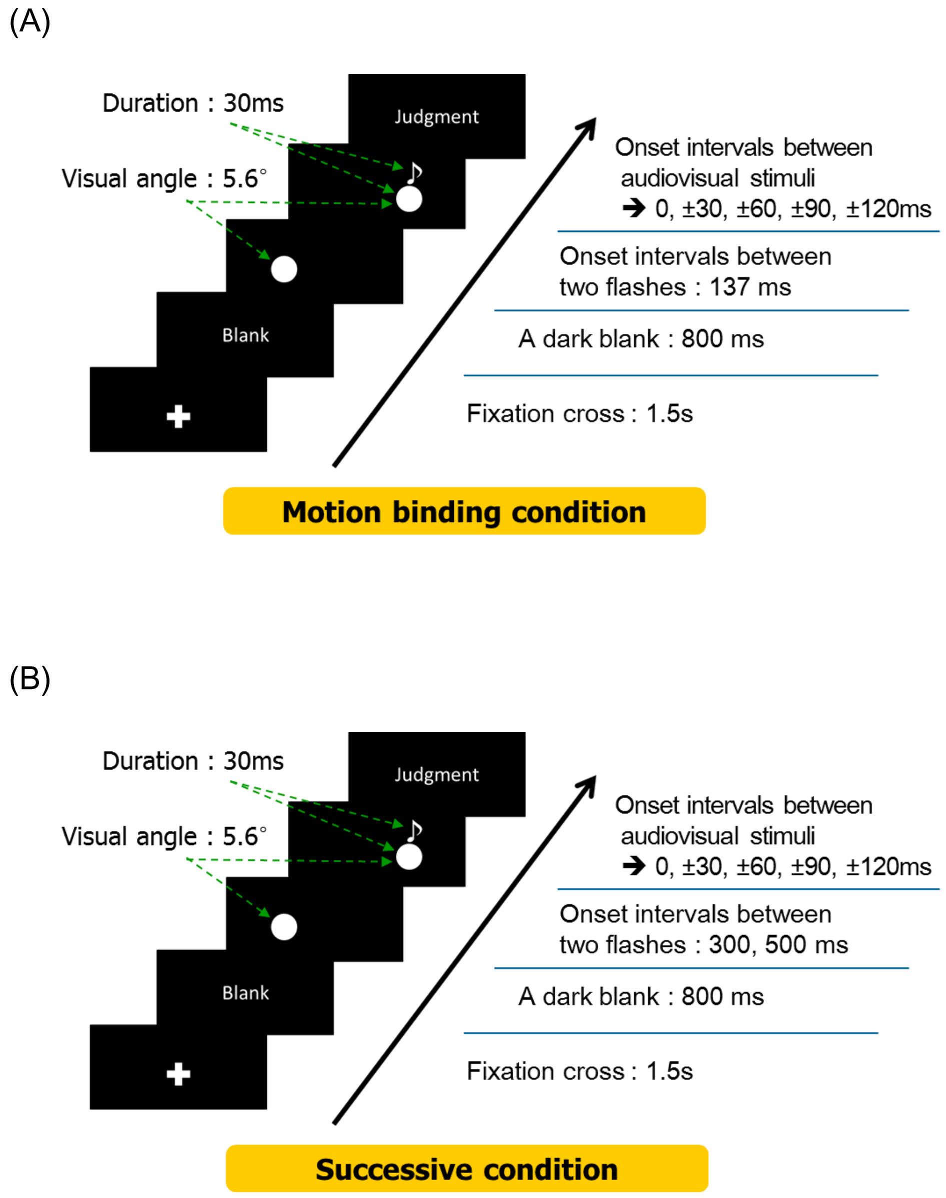


Figure 3.9. Schematic illustration of Experiment 2. (A) and (B) show the procedure for Section 1. After the presentation of a fixation cross (1.5 s) and a blank screen (800 ms), two visual stimuli were presented in the motion binding condition (A) and two visual stimuli with SOAs of 300 and 500 ms in the successive condition (B). Sounds were presented either before or after the visual stimuli at SOAs ranging from –120 to 120 ms at 30 ms intervals in random order.

In Experiment 3, the procedure was the same as in Chapter 2 with the following exceptions. Experiment 3 consisted of two kinds of TOJ tasks in the motion binding condition (Figure 3.10 (A)) and the non-motion binding condition (Figure 3.10 (B)). The motion binding condition was equivalent to that of Chapter 2, with an SOA of 137 ms between the two flashes. The non-motion binding condition consisted of SOAs of 40 or 300 ms between the two flashes. The timing of the auditory stimulus relative to the second flash was the same as in Chapter 2. The participants were instructed to judge the order of the second visual stimulus and the brief sound. Experiment 3 consisted of 405 trials (3 visual conditions × 9 audiovisual SOAs × 15 repetitions) in counterbalanced order and the experiment was divided into nine blocks of 45 trials (9 audiovisual SOAs × 5 repetitions). Including practice for each task, each experiment took approximately one and a half hours.

Before starting each experiment, we examined whether the participants perceived motion between two flashes. In Experiments 1, 2 and 3, we showed the participants two flashes with an SOA of 137 ms 10 times, and participants were asked to evaluate whether their impression of the stimuli was of "continuous motion" or "discrete or non-motion stimuli." In Experiments 1, 2 and 3, we showed them two flashes, with SOAs of 137 ms, 40 ms and 300 ms, 10 times in counterbalanced order. Then, the participants were asked to evaluate whether their impression of the stimuli was of "continuous motion" or "discrete or non-motion stimuli." Only the participants who perceived "continuous motion" with an SOA of 137 ms and "discrete or non-motion stimuli stimuli" with SOAs of 40 ms and 300 ms for all stimuli continued to participate in the experiment. We confirmed that motion was perceived during the TOJ task after each experimental session was completed.

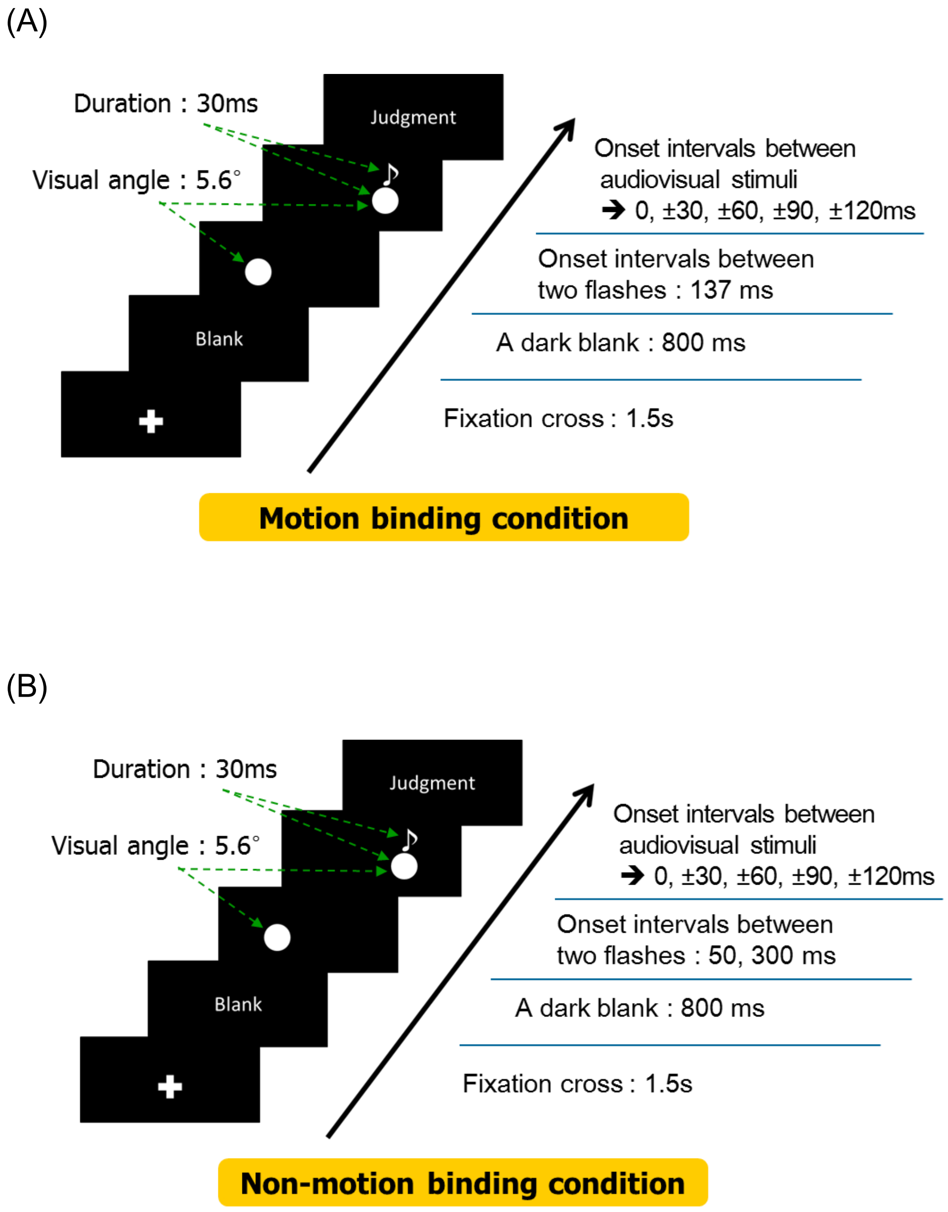


Figure 3.10. Schematic illustration of Experiment 3. (A) and (B) show the procedure for Experiment 3. After the presentation of a fixation cross (1.5 s) and a blank screen (800 ms), two visual stimuli were presented in the motion binding condition (A) and two visual stimuli with SOAs of 40 and 300 ms in the non-motion binding condition (B). Sounds were presented either before or after the visual stimuli at SOAs ranging from –120 to 120 ms at 30 ms intervals in random order.

3.3.2.5 Data Analysis

In Section 3, the Data analysis was the same as in Chapter 2.

3.3.3 Results

In Chapter 3.3, the participants performed three kinds of Experiments in the motion binding and non-motion binding conditions. Individual PSSs and JNDs for the two conditions were computed as in Chapter 2. In Experiment 1, Figure 3.11 (A) illustrates psychometric curves fitted to the distribution of the mean TOJ data for all participants in the motion binding and simultaneous conditions. As shown in Figure 3.11 (B), the PSS in the simultaneous condition had a positive value of 14.39 ms (SE = 6.06), but the PSS in the motion binding condition shifted toward a sound-lead stimulus of –3.57 ms (SE = 5.35). A paired t-test of PSSs indicated a significant difference between the TOJ tasks in the motion binding condition and those in the simultaneous condition (t(11) = –2.43, *p* = 0.034, see Table A-7). In addition, the JND in the motion binding condition was smaller than that in the simultaneous condition (see Figure 3.11 (C)), and the JND values were 26.42 ms (SE = 4.59) and 39.48 ms (SE = 6.47), respectively. A significant difference between the JNDs was observed in the paired t-test (t(11) = –4.49, *p* = 0.0009, see Table A-8).

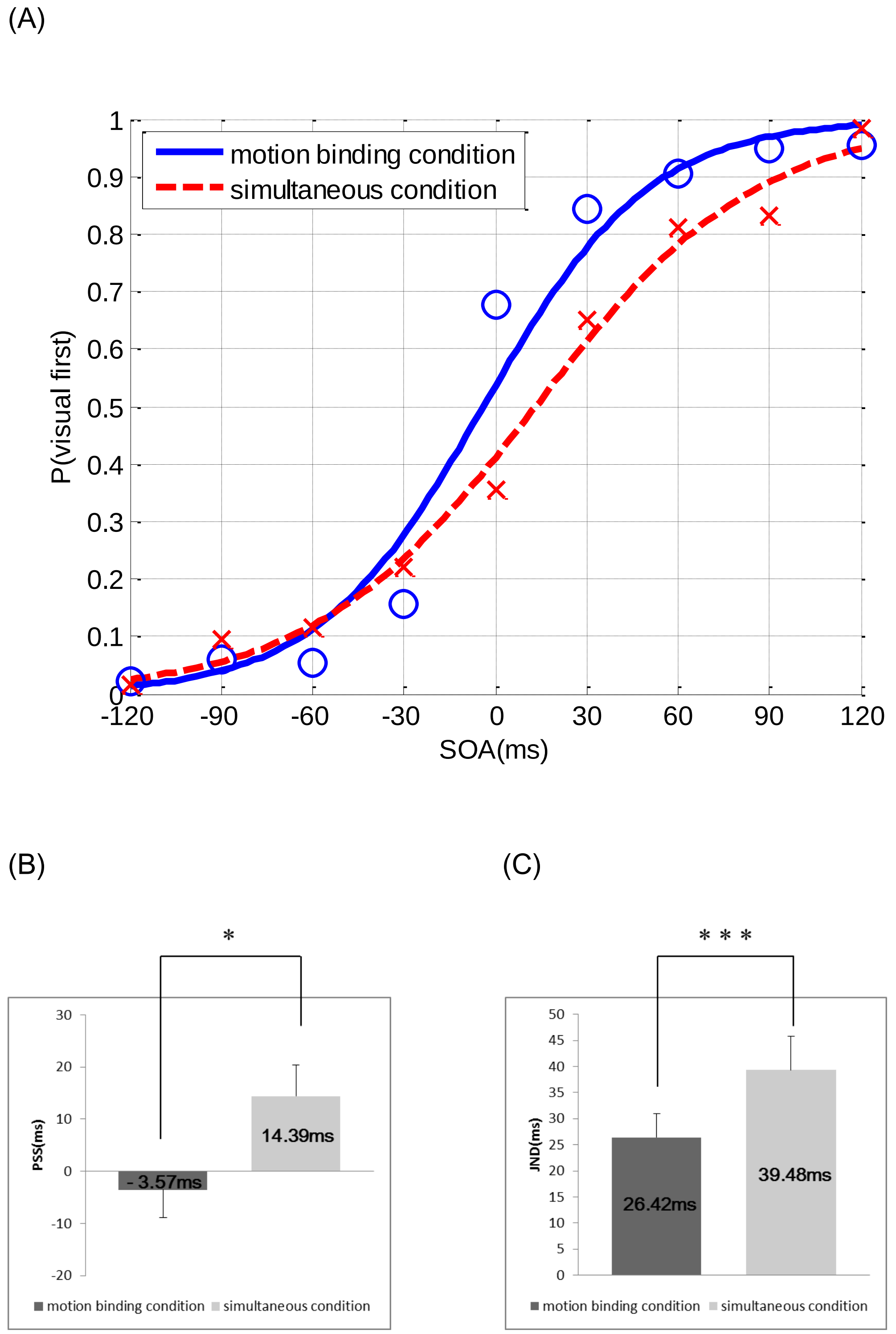
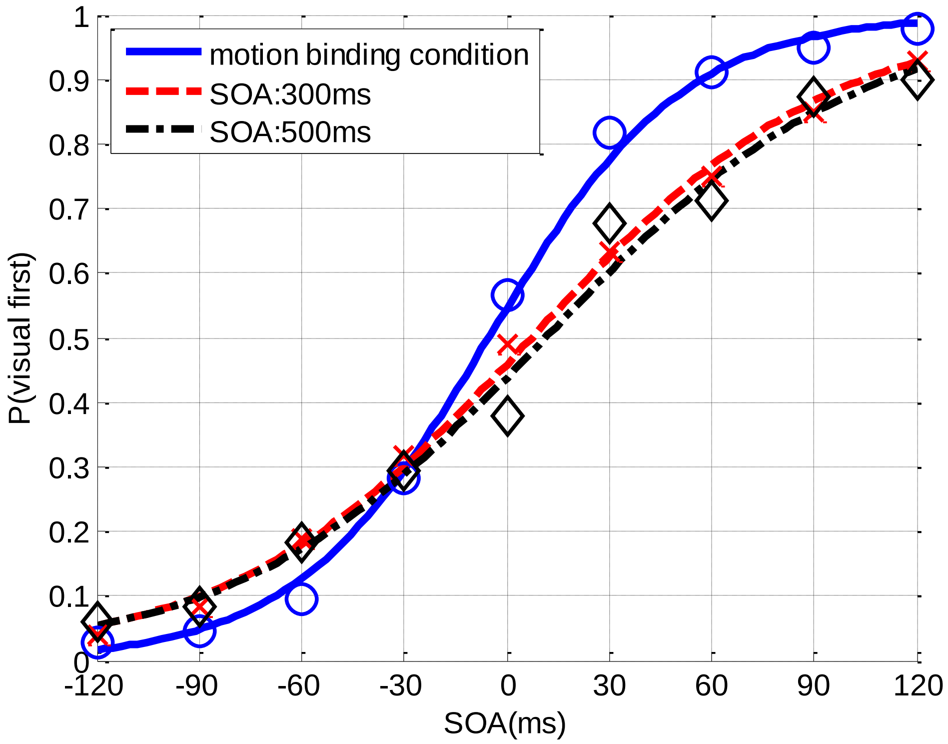


Figure 3.11. Results of Experiment 1. (A) shows psychometric curves fitted to the distribution of mean TOJ data for all participants in the motion binding condition and in the normal condition. Horizontal axis shows the onset time of the auditory

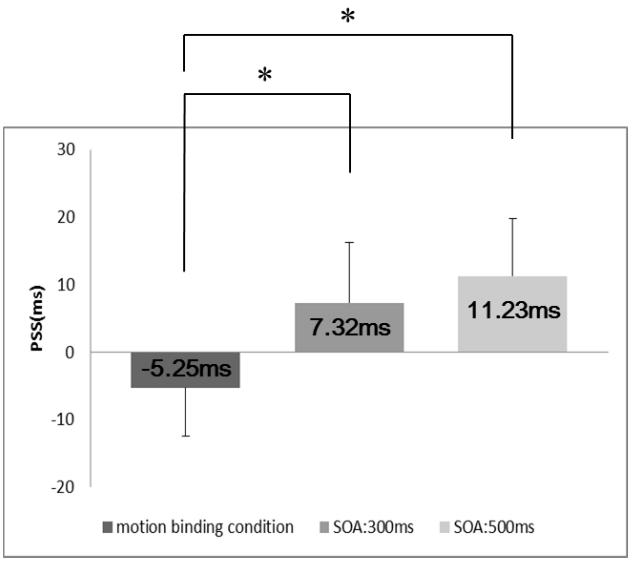
stimulus paired with a visual stimulus. The negative values of the SOAs indicate that an auditory stimulus preceded a visual stimulus. (B) shows the mean PSSs in the motion binding and normal conditions. Error bars represent the standard error of the means. (C) shows the mean JNDs in the motion binding and normal conditions. Error bars represent the standard error of the means, *: $p < .05$, **: $p < .01$, ***: $p < .001$, for a paired t-test.

In Experiment 2, the participants performed both kinds of TOJ tasks in the motion binding and successive conditions. Individual PSSs and JNDs for the two conditions were computed as in Chapter 2. Figure 3.12 (A) illustrates the results of Experiment 2 with psychometric curves fitted to the distribution of the mean TOJ data of all participants in the motion binding and successive conditions. Figures 3.12 (B) and (C) show the results of PSS and JND in Experiment 2. The PSS shifted toward a sound-lead stimulus in the motion binding condition, and the JND in the motion binding condition was smaller than that in the successive condition. A repeated-measures analysis of variance (ANOVA) of the PSSs showed a significant main effect for the temporal interval, $F(2, 23) = 15.83$, $p < 0.001$ (see Table A-9). Multiple comparisons with Holm correction showed significant differences between the motion binding condition and the successive condition (motion binding and an SOA of 300 ms: $p = 0.042$, motion binding and an SOA of 500 ms: $p = 0.012$, see Table A-9). Moreover, a repeated-measures ANOVA of the JNDs revealed a significant main effect of the temporal interval, $F(2, 23) = 25.03$, $p < 0.001$ (see Table A-10). In addition, multiple comparisons with Holm correction confirmed that there was a significant difference between the motion binding and successive conditions (motion binding and an SOA of 300 ms: $p = 0.002$, motion binding and an SOA of 500 ms: $p < 0.001$, see Table A-10).

(A)


motion binding condition
SOA:300ms
SOA:500ms
P(visual first)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-120
-90
-60
-30
0
30
60
90
120
SOA(ms)

(B)


*
*
PSS(ms)
30
20
10
0
-10
-20
-5.25ms
7.32ms
11.23ms
motion binding condition
SOA:300ms
SOA:500ms

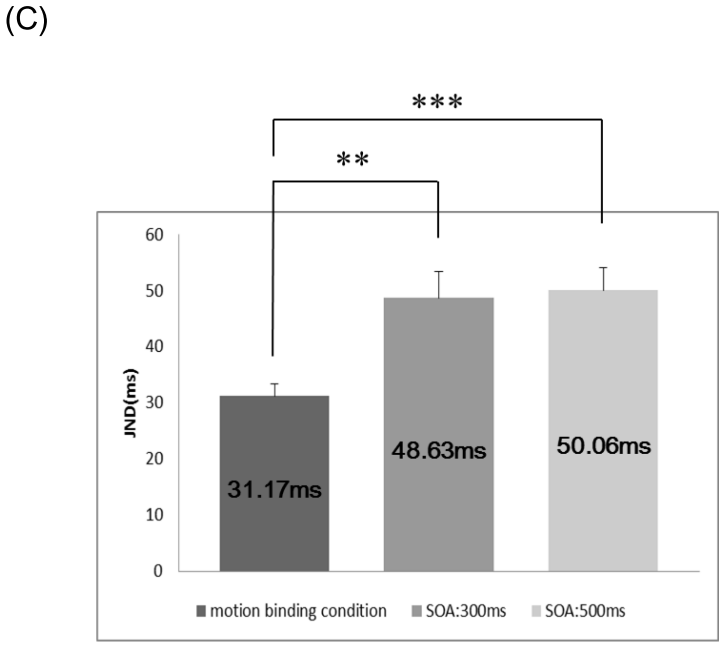
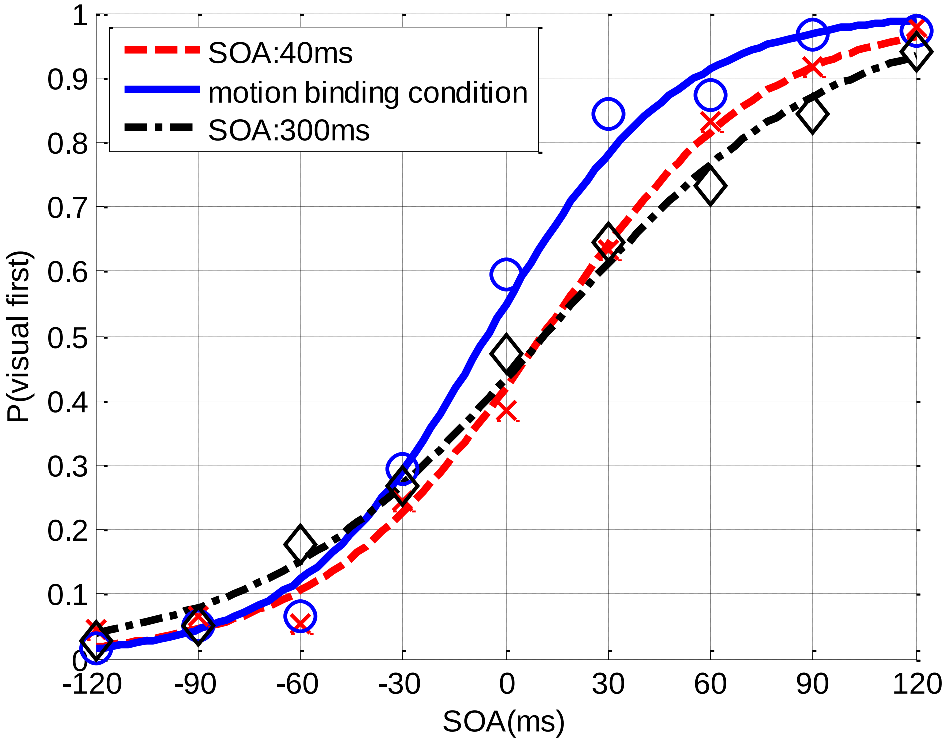


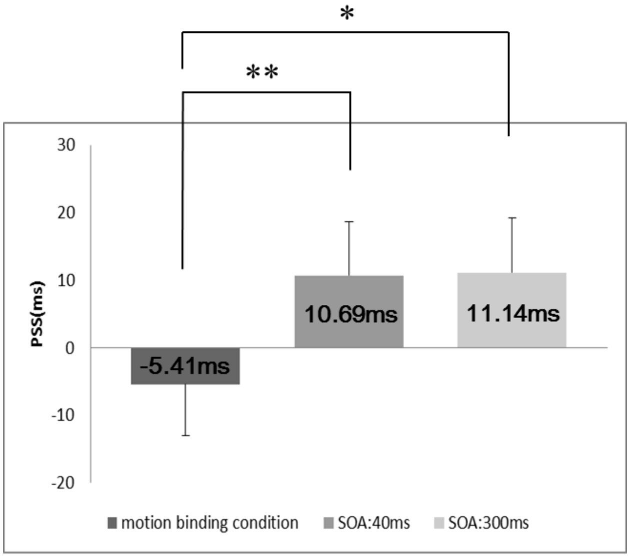
Figure 3.12. Results of Experiment 2. (A) shows psychometric curves fitted to the distribution of mean TOJ data for all participants in the motion binding and successive conditions. The negative values of the SOAs indicate that an auditory stimulus preceded a visual stimulus. (B) shows the mean PSSs in the motion binding and successive conditions. Error bars represent the standard error of the means. (C) shows the mean JNDs in the motion binding and successive conditions (SOA: 300 ms and 500 ms). Error bars represent the standard error of the means. The p values were adjusted for multiple comparisons using Holm’s correction, *: p < .05, **: p < .01, ***: p < .001.

Figure 3.13 (A) illustrates the results of Experiment 3 with psychometric curves fitted to the distribution of the mean TOJ data of all participants in the motion binding and non-motion binding conditions. Figures 3.13 (B) and (C) show the results of PSS and JND in Experiment 3. The PSS shifted toward a sound-lead stimulus in the motion binding condition, and the JND in the motion binding condition was smaller than that in the non-motion binding condition. A repeated-measures analysis of variance (ANOVA) of the PSSs showed a significant main effect for the temporal interval, $F(2, 22) = 8.473$, $p < 0.002$ (see Table A-11). Multiple comparisons with Bonferroni correction showed significant differences between the motion binding condition and the non-motion binding condition (motion binding and an SOA of 40 ms: $p = 0.009$, motion binding and an SOA of 300 ms: $p = 0.031$, see Table A-11). Moreover, a repeated-measures ANOVA of the JNDs revealed a significant main effect of the temporal interval, $F(2, 22) = 6.053$, $p < 0.008$ (see Table A-12). In addition, multiple comparisons with Bonferroni correction confirmed that there was a significant difference between the motion binding and non-motion binding conditions (motion binding and an SOA of 300 ms: $p = 0.002$, see Table A-12), but there was no significant difference between the motion binding and an SOA of 40 ms: $p = 0.634$ (see Table A-12).

(A)


SOA:40ms
motion binding condition
SOA:300ms
P(visual first)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
-120
-90
-60
-30
0
30
60
90
120
SOA(ms)

(B)


*
**
30
20
10
0
-10
-20
PSS(ms)
-5.41ms
10.69ms
11.14ms
motion binding condition
SOA:40ms
SOA:300ms

(C)

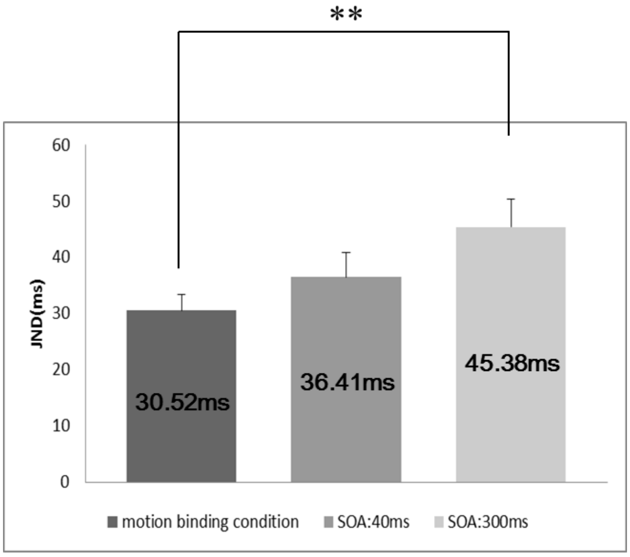


Figure 3.13. Results of Experiment 3. (A) shows psychometric curves fitted to the distribution of mean TOJ data for all participants in the motion binding and non-motion binding conditions. The negative values of the SOAs indicate that an auditory stimulus preceded a visual stimulus. (B) shows the mean PSSs in the motion binding and non-motion binding conditions. Error bars represent the standard error of the means. (C) shows the mean JNDs in the motion binding and non-motion binding (SOA: 40 ms and 300 ms). Error bars represent the standard error of the means. The p values were adjusted for multiple comparisons using Bonferroni's correction, *: p < .05, **: p < .01, ***: p < .001.

3.3.4 Discussion

The PSS shifted toward a sound-lead stimulus and especially became closer to physical simultaneity (i.e., zero) in the motion binding condition, and the JND in the motion binding condition was smaller than that in the simultaneous, successive and non-motion binding condition. This shows that motion binding property shifted simultaneity toward physical simultaneity, and reduces the temporal window in audiovisual integration. In previous studies, the characteristics of perceptual binding in cross-modality have been reported. For example, unity assumption well represents the multi-modal binding, which means that different sensory information is bound and perceived as a same event (Bedford, 2001; Radeau and Bertelson, 1977; Bertelson and Aschersleben, 1998; Vatakis and Spence, 2007; Welch, 1999; Welch and Warren, 1980). Also, it has been reported that one sensory information is captured by another sensory information (Gebhard and Mowbray, 1959; Fendrich and Corballis, 2001; Bertelson and Aschersleben, 2003; Morein-Zamir, et al., 2003; Freeman and Driver, 2008; Recanzone, 2003; Vroomen and de Gelder, 2004). For example, temporal ventriloquism effect, if there is temporal separation in audiovisual stimulation, temporal events of the visual stimuli are bound by that of auditory stimuli (Bertelson and Aschersleben, 2003; Morein-Zamir, et al., 2003). However, there were no studies that have investigated the effect of motion binding property on the multisensory binding. In this study, we investigated the effect of motion binding property on audiovisual simultaneity. As a results, motion binding property affected audiovisual simultaneity and contributes to accurate perceptions of temporal events.

In the Experiment 3, although the PSS in the motion binding condition differed from that in the non-motion binding condition, there was no difference in JNDs between the motion binding condition and the SOA of 40 ms in the non-motion binding condition. With respect to visual attention, it has been reported that there are two types of visual attention such as exogenous attention and endogenous attention. Exogenous attention is an involuntary system, which rises and decays quickly and acts on maximally about 100-120 ms and is effective up to 300 ms (Muller and Rabbitt, 1989; Hein, et al., 2006; Ling and Carrasco, 2006; Liu, et al., 2007; Nakayama and Mackeben, 1989; Remington, et al., 1992). On the other hand, endogenous attention is a voluntary system that deployed at more than 300 ms (Muller and Rabbitt, 1989; Hein, et al., 2006; Ling and Carrasco, 2006; Liu, et al., 2007; Nakayama and Mackeben, 1989; Remington, et al., 1992). In the results of Experiment 3, there is a possibility that exogenous attention by a short temporal interval is activated in motion binding perception and endogenous attention by long temporal intervals is allocated in the successive perception. Therefore, there was a need to compare between motion binding perception and non-motion binding perception by a short temporal interval to confirm the influence of exogenous attention. As a result, exogenous attention by a short temporal interval (the SOA of 40 ms in the non-motion binding condition) resulted in small JND, but unaffected the PSS. This means that exogenous attention reduces the temporal window in audiovisual integration but does not affect the temporal orders for simultaneity. From this, motion binding property affects audiovisual temporal orders for simultaneity and audiovisual temporal orders are more accurately integrated during motion perception. Therefore, it revealed that motion binding property contributes to the accurate perceptions of temporal orders in audiovisual processing.

Our findings reveal that visual apparent motion affects audiovisual simultaneity by shifting simultaneity toward physical simultaneity, and reduces the temporal window in audiovisual integration. This suggests that visual motion information contributes to accurate perceptions of temporal events in the physical world. In particular, we suggest that motion binding property affects accurate perceptions of temporal simultaneity in multisensory processing. This binding property will prompt developmental research on motion perception and multisensory integration.

Chapter 4

Motion on Simultaneity in Human Communication

4.1 Introduction

The influence of visual motion information on audiovisual simultaneity have confirmed in Chapters 2 and 3. As a result, visual motion information shifted audiovisual simultaneity toward physical simultaneity, and reduces the temporal window in audiovisual integration. This suggests that visual motion information contributes to accurate perceptions of temporal events in the physical world. In particular, the findings reveal that motion binding property in visual perception is the mechanism affecting accurate perceptions of temporal simultaneity in multisensory processing.

However, we are exposed to an interpersonal interaction situation (see Figure 4.1). In human-human interaction, two or more people transmit and perceive multisensory information each other. In particular, we integrate the multisensory signals by simultaneity perception and the simultaneity perceptions interact between two individuals. Human communication (or interpersonal communication) is defined as this process for information exchange between two or more people. Human communication consists of verbal communication and nonverbal communication through mainly auditory and visual channels. Interestingly, it is reported that body movements between two or more people are synchronized during communication process. This phenomenon is called body movement synchronization, which is achieved by the interaction of simultaneity perceptions between two or more individuals. Chapter 4 is aimed to confirm the influence of motion information on simultaneity perceptions (using body movement synchronization) in human communication.

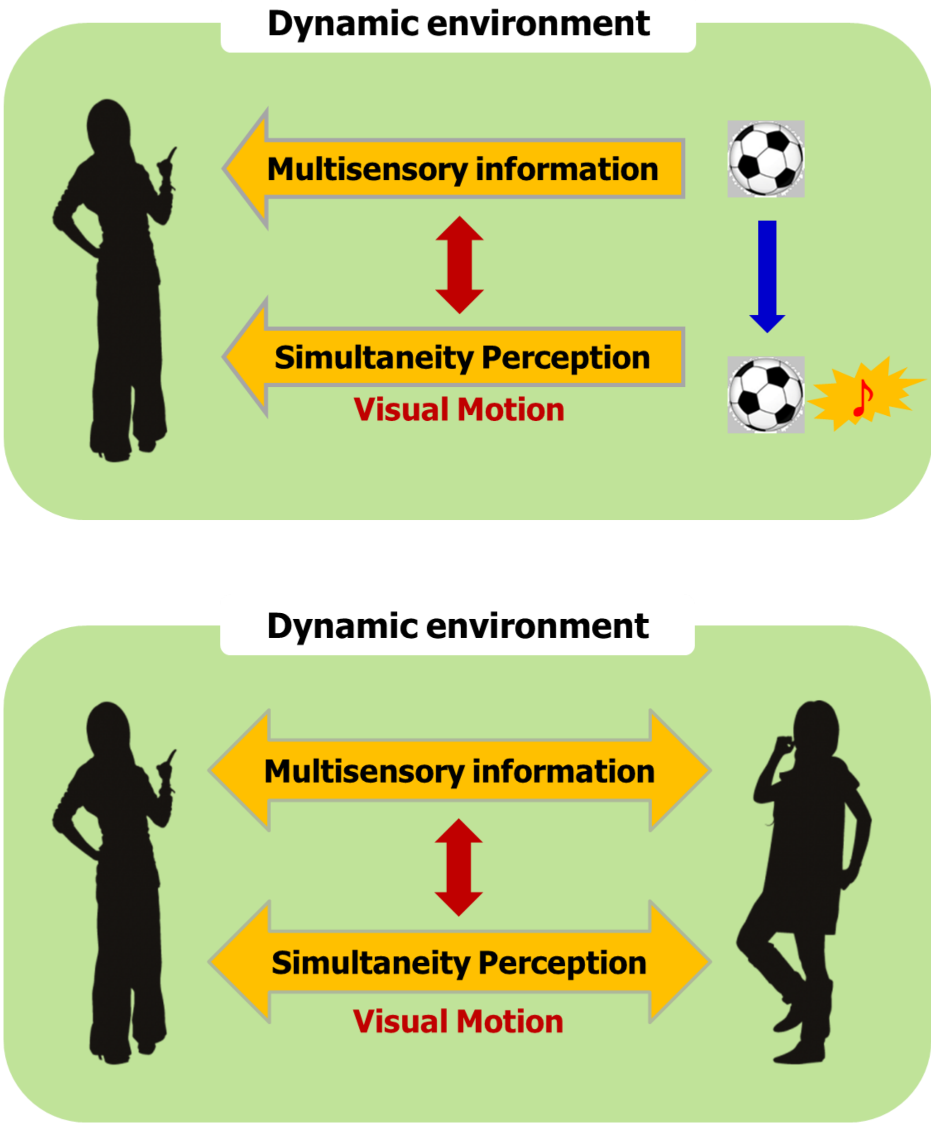


Figure 4.1 Human-environment and human-human interactions. We interact with environment by perceiving and interpreting outer world. In human-human interaction, two or more people transmit and perceive multisensory information each other and the multisensory signals interact between two persons. In particular, we integrate the multisensory signals by simultaneity perception and also the simultaneity perceptions interact between individuals.

4.1.1 Simultaneity in Various Perceptual Scales

Simultaneity perception has been shown in a variety of perceptual scales. For example, the temporal window of integration in audiovisual speech was larger than that in an experiment using a simple pair of sound and flash (see Table 2-1) (Zampini, et al., 2003; Wassenhove, et al., 2007). This means that conversation attributes in audiovisual speech change the temporal window of audiovisual integration.

The characteristics of simultaneity perception has been shown in the relations between sensory perception and body movement. For example, the body movement affects sensory perception by altering simultaneity perception (Frissen, et al., 2010; Nishi, et al., 2014). In particular, there is a phenomenon called ‘negative asynchrony’, which body motion occasionally precedes sensory perception (Miyake, et al., 2004; Miyake, et al., 2007; Repp, 2005; Repp, 2008; Woodrow, 1932; MacDorman, 1962). Also, mirror neuron system has been reported that mirror neurons are particular neurons that fire when an animal does an action and when it observes a similar action either. (Rizzolatti and Craighero; 2004; Oztop, et al., 2012; Arbib, 2010; Buccino, et al., 2004; Rizzolatti, 2005; Iacoboni and Dapretto, 2006). These researches tell us the interaction between sensory perception and body motion, and they are an ‘inseparable’ relation. In particular, body motion is deeply involved in sensory perception and may contribute to the simultaneity perception close to real time by changing the sensory perception.

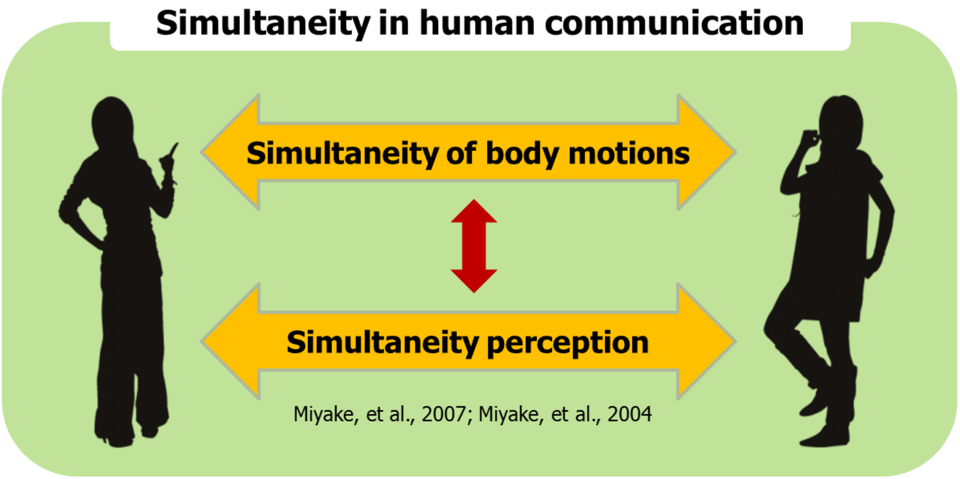


Figure 4.2. Relation between simultaneity perception and simultaneity of body motions. 'Simultaneity' has been shown in a variety of perceptual scales. The characteristics of simultaneity has been studied in the relations between sensory perception and body motion. Many researches have reported the interaction between sensory perception and body motion, and they are an 'inseparable' relation. Many studies have reported the interaction between sensory perception and body motion, and they are an 'inseparable' relation. In particular, body motion is deeply involved in simultaneity perception and they interact with each other.

Through the interaction between perception and body motion, we can communicate with each other smoothly. In human communication, verbal and nonverbal information is transmitted through the auditory and visual channels including the various meaning for communication (Richmond et al, 2008). Therefore, we communicate with each other by perceiving and integrating the audiovisual information simultaneously through verbal and nonverbal information, and human communication shows the most high-order and complex processing in sensory integration. Interestingly, body movement synchronization as simultaneity between two or multiple people has been shown in human communication. The most important point is that these body movement synchrony is based on simultaneity perception, which is a process perceiving body movements of each other as visual motion (see Figure 4.2). Therefore, body movement synchronization is achieved by the interaction of simultaneity perceptions between two or more individuals.

4.1.2 Simultaneity in Human Communication

As mentioned above, human communication consists of verbal and nonverbal communications (see Figure 4.3). In particular, we know that verbal communication occurs with awareness, whereas nonverbal communication occurs subconsciously or automatically (Richmond, et al., 2008; Schneider and Chein, 2003; Bargh and Chartrand, 1999; Bargh, 1989; Lakin, 2006). We are unable to control nonverbal communication consciously, and therefore nonverbal responses are known to be immediate and honest (Altmann, et al., 2007). Nonverbal communication contributes to human communication in a variety of ways and is an important factor in social interaction (Richmond, et al., 2008; Hodgins and Zuckerman, 1990; Tickle-Degnen and Rosenthal, 1990). Among nonverbal behaviors that contribute to human communication, nodding the head is known to be a universal behavior in many cultures, and it plays an important role as a form of feedback in human communication (Darwin, et al., 1998; Kita and Ide, 2007; Maynard, 1987).

Of specific relevance to this study, body movement synchronization is known to be a dominant characteristic in nonverbal communication. In face-to-face communication, body movement synchronization has been observed in social and interpersonal relationships. For example, the body movements of neonates synchronize with the speech of their mothers (Condon and Sander, 1974), and intimate partners frequently interact by synchronizing their postures and body movements (Bernieri, et al., 1996). In particular, Hove and Risen (2009) concluded that interpersonal synchrony is the critical factor contributing to likability with an increase of affiliation and Marsh et al. (2009) have reported that behavioral and embodied methods can be used for investigating the relationship between sociality and coordination with other individuals, which is fundamental and serves as the basis for our social connection to others. In addition, the strong synchronization of body movements between clients and their psychotherapy counselors has been found for positively evaluated counseling groups (Komori and Nagaoka, 2010). Specifically, head nods often occur at the same time, even during conversations among multiple participants (Kita and Ide, 2007). Some researchers report that the synchronization of head nods reflects positive emotions in interpersonal relationships (Kita and Ide, 2007; Maynard, 1987).

With respect to head nod, Maynard (1987) proposed that head movements may be categorized with discourse context. With regard to listener, head movement has a function as back-channel continuers. This head movements occur with or without brief verbalizations such as ‘uh huh’ and ‘right’, etc. Second function is a turn-shift or turn-taking. Head movement occurs when turn shifts and when the listener takes a new turn. On the other hand, the speaker’s head movements occur with the last syllable of the speech. In addition, the speaker uses head movement during the turn-transition period and with the sentence final word or syllable. Also, head movement is used by the speaker with affirmation and agreement such as ‘yes’, ‘yeah’ and ‘that’s right’. Among others head movement occurs in the context where a yes/no answer is expected. Therefore, the functions of head movement in listener can be classified into two, which are back-channel continuers and turn-taking.

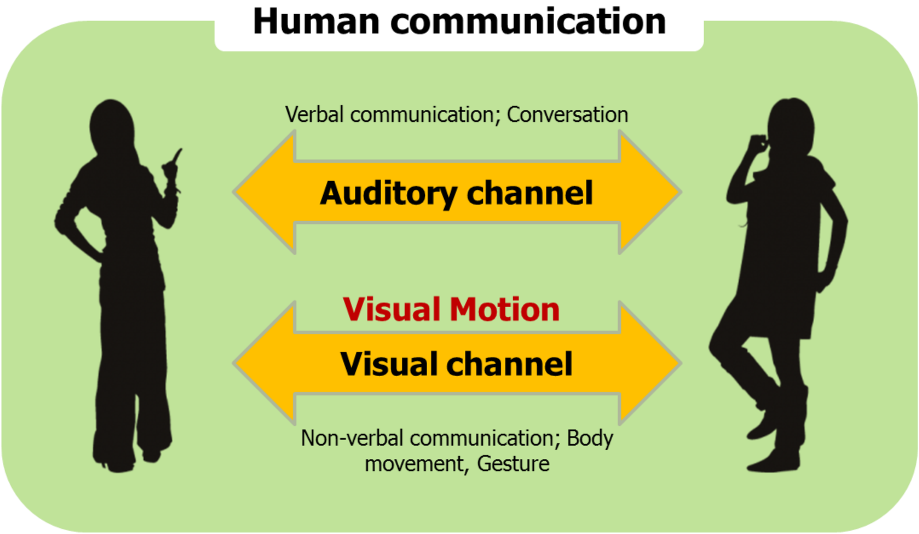


Figure 4.3. Structure of human communication. Human communication consists of verbal and nonverbal communications. In human communication, verbal and nonverbal information is transmitted through the auditory and visual channels including the various meaning for communication (Richmond et al, 2008). Therefore, we communicate with each other by perceiving and integrating the audiovisual information simultaneously through verbal and nonverbal information, and human communication shows the most high-order and complex processing in sensory integration.

4.1.3 Newly-devised Detection Method

With respect to the detection of body movement synchrony, Bernieri (1988a) and Bernieri et al. (1988b) have analyzed body movement synchrony by observer ratings through movement synchrony perceived in video clips. Also, previous researchers have used video-based analysis to measure body movement synchronization (Kita and Ide, 2007; Maynard, 1987; Condon and Sander, 1974; Bernieri, et al., 1996; Komori and Nagaoka, 2010; Nagaoka and Komori, 2008; Ramseyer and Tschacher, 2011). In particular, some researchers have analyzed the synchronization of body movements by focusing on changes in the amplitude of unspecified body movements with a predetermined video frame rate (Komori and Nagaoka, 2010; Nagaoka and Komori, 2008; Ramseyer and Tschacher, 2011; Paxton and Dale, 2013). However, there are still three problems with the analysis of synchronized body movements. First, there is no research on phase difference as an indicator of body movement synchronization in human communication. In theoretical studies, phase difference is a very important factor in synchronization analysis in rhythmic and periodic motion because it shows the most accurate temporal relationship in synchronization (Kuramoto, 1984; Acebron, et al., 2005; Schmidt, et al., 1990). Therefore, it is necessary to measure phase differences in body movement synchronization and to define nonverbal synchronization quantitatively according to the distribution of phase differences in human communication. Second, many previous studies did not consider specific interactions between nonverbal behaviors in synchronization analysis because they have comprehensively analyzed the overall and unspecified body movements. However, specifying the body part concerned is essential for studying nonverbal synchronization (Paxton and Dale, 2013). Third, the video frame rate (10 or 30 f/s) used in previous studies was insufficient to analyze the details of body movements (Komori and Nagaoka, 2010; Nagaoka and Komori, 2008; Ramseyer and Tschacher, 2011). Therefore, we need to acquire time-series data with high temporal resolution by attaching a measuring device directly to participants' bodies.

The purpose of this study was to investigate the influence of visual motion on simultaneity perception in human communication. Also, this study provides a quantitative definition of phase difference and a new method using the phase difference distribution for detecting body movement synchronization in human communication. We characterized body movement synchronization using four statistical measurements of the phase difference distribution. These four measurements include: density as an indicator of the synchronization activity (frequency of synchronization), mean phase difference as an indicator of the synchronization direction (tendency of leader–follower relationship), and standard deviation (SD) and kurtosis as indicators of the synchronization strength (stability of the tendency).

To confirm the influence of visual motion, we focused on differences in body movement synchronization under different types of communication situations: direct face-to-face communication and remote communication via television. Also, we applied our new method to human communication in which the roles of speaker and listener were defined to confirm the validity of our definition. Most importantly, body movements are coordinated between perceptually coupled individuals (Schmidt, et al., 1990). Furthermore, Bernieri (1988a) and Bernieri et al. (1988b) have emphasized the importance of interpersonal interaction in body movement synchrony by comparing genuine synchrony with true interaction and pseudo-synchrony with no interaction. Therefore, in this study, we examined the difference in phase difference distribution between face-to-face communication with visual interaction and remote communication with unidirectional visual perception. The face-to-face communication condition is set up as a situation in which two participants are visually coupled, whereas the remote communication condition is set up as a situation in which two participants are not visually coordinated in which the listener has visual information about the speaker but the speaker has no visual access to the listener. In other words, a pair of participants can bilaterally perceive the visual information in direct face-to-face communication. However, in remote communication, the listener only perceives visual information unilaterally.

Moreover, technologies developed in recent years enable us to communicate remotely as well as in person (van der Kleij, et al., 2005; van der Kleij, et al., 2009).

Although remote communication has been developed to approximate face-to-face communication, remote communication, especially via television, remains inadequate compared with face-to-face communication (Harboe, et al., 2008; Cesar, et al., 2008). Therefore, there is a need to investigate the differences between face-to-face and remote communications.

From two experiments, we examined the influence of visual motion on simultaneity perception through body movement synchronization. In the face-to-face communication condition, participants performed a lecture task via face-to-face communication. We examined phase differences in time-series data on the acceleration of head nods between two participants throughout the lecture task, and we detected the synchronization of head nods from the distribution of the phase differences. In the remote communication condition, participants performed the lecture task remotely via television, and again we detected the synchronization of head nods from the distribution of the phase differences in time-series data on the acceleration of head nods between pairs of participants.

4.2 Materials and Methods

4.2.1 Experimental Design

We used a lecture task in this study to distinguish clearly between the speaker and listener during the communication process. By having the participants perform the lecture task twice in the face-to-face communication and remote communication conditions, we allowed the listener to adapt to the task and to predict the content of the task by the learning effect. Therefore, pairs of participants were divided into two groups, and they performed the lecture task separately. In the face-to-face communication condition, a teacher who takes over the role of speaker delivered certain content to a student who takes over the role of listener in face-to-face communication. In the remote communication condition, a pair of participants performed the lecture task remotely (in different rooms) via television. The size of the teacher's face, the volume of the teacher's voice, and the gaze point between the teacher and student were identical in the face-to-face communication and remote communication conditions. The listener was only allowed back-channel signals during the lecture task in the face-to-face communication and remote communication conditions, and the constraints for the lecture task were the same in both experiments.

4.2.2 Participants

Twelve pairs of subjects (16 males and eight females, all in their 20s) participated in the face-to-face communication and remote communication conditions. We derived the following selection criteria for pairs of participants from a previous study (Maynard, 1987; Maynard, 1990): the partners should differ in age by less than five years, be of the same sex, and be native speakers of Japanese. In addition, we imposed the condition that only two people would interact with each other during the experiment. The ethics committee of the Tokyo Institute of Technology specifically approved this study, and written informed consent was obtained from each participant to participate in this study.

4.2.3 Apparatus

We used a small three-axis acceleration sensor (4.5 cm × 4.0 cm) with a sampling frequency of 100 Hz (WAA-006, Wireless Technologies, Japan) to measure time-series data on the acceleration of head nods. The data were recorded on a PC (Latitude E5400, Dell, TX, USA) via Bluetooth. The acceleration sensor was attached to the forehead of each participant (see Figure 4.4 (A)). In addition, we used three video cameras (Xacti, Sanyo, Japan) to record the overall situation of the teacher and student participants. In the remote communication condition, a video camera (HDR-CX270, Sony, Japan) in the teacher's room recorded images of the teacher and transmitted them to a television (60-inch LED display, with 1920 × 1080 pixel resolution, UN60ES8000F, Samsung, Korea) in the student's room. The video camera and television were connected by an HDMI cable, and another camera (Xacti, SANYO, Japan) recorded the student.

4.2.4 Procedures

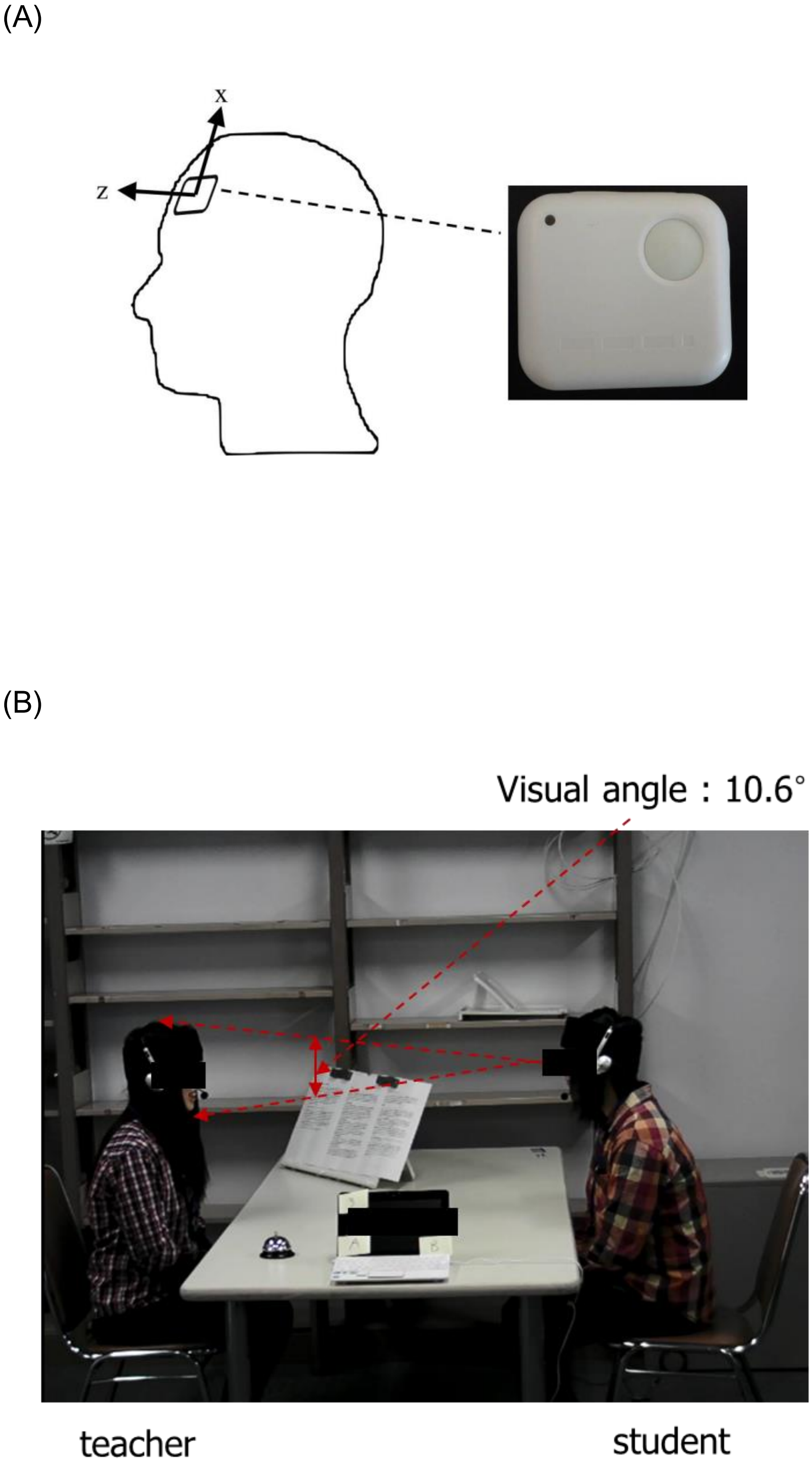
In the face-to-face communication condition, each participant was randomly assigned to the role of either teacher or student, and before the experiment began, the teacher was given a Wikipedia article. The article was a “cold reading,” related to the techniques of persuasion (Cold reading). The criteria for selecting the article were that it should be on a less well-known topic and that it would take approximately 5–10 minutes to describe. The article was three A4 pages and 2,759 Japanese characters in length. The teacher was separated from the student and instructed to read the article to understand the content. The teacher then summarized the article freely for 5 to 10 minutes. The teacher removed unnecessary content from the summary and then practiced describing the article to the experimenter and in his/her own words. At the start of the experiment, the teacher sat face-to-face with the student across a table at a distance of 1.2 meters, with a visual angle of 10.6° for the teacher’s face (see Figure 4.4 (B)). The temperature of the room was 24.2°C, the illuminance was 913.8 lux (CL-200A, Konica Minolta, Japan) and environmental noise was 34.3 dB (AR814, Smart Sensor, China). The article was placed on a book stand in front of the teacher, who described the article to the student in approximately 5 to 10 minutes in Japanese. The teacher was instructed to speak in a loud and clear voice, and to look the student in the eye while speaking. The student was asked to look the teacher in the eye, to listen carefully to the teacher’s description, and to learn the content. The students were not allowed to ask questions; they were only allowed to use back-channel signals, including head nods and short utterances such as “un,” “hai” and “ee,” which are equivalent to “mmhm,” “uh huh” and “yeah” in English (Kita and Ide, 2007; Maynard, 1987; , Yngve, 1970; White, 1989; Angles, et al., 2000).

We instructed the participants to minimize the influence of body movements except head nods, and we imposed the following constraints to help them do so.

- The teacher was not allowed to show the manuscript to the student.

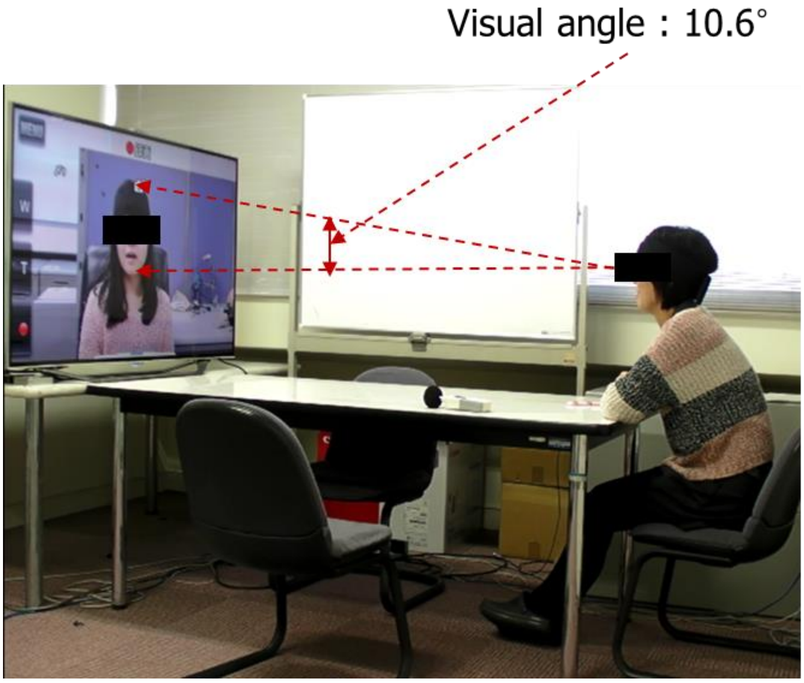
- Neither teacher nor student could change posture significantly.
- Neither teacher nor student could touch the sensor during the experiment.

We used the same procedure in the remote communication condition as in the face-to-face communication condition, except for the following points. The teacher and student sat in separate rooms, and the lecture was given via television. The teacher sat in front of a video camera and described the same article as in the face-to-face communication condition for approximately 5 to 10 minutes. There was no difference in the length of the description time between the face-to-face communication and remote communication conditions (t (22) = 0.057, P = 0.955, see Tables A–13 and A–14). During the practice, we measured the sound level of the teacher's description every 10 seconds with a digital sound level meter (AR814, Smart Sensor, China), and the volume of the television was adjusted to the actual range of the volume of the teacher's voice (Mean: 62.3 dB; SD: 5.4 dB). During the experiment, the teacher was asked to look at the camera while speaking, as if speaking face-to-face with the student. The audiovisual information of the teacher was transmitted to a television in the student's room via a video camera. The student sat in front of the television at a distance of 1.8 meters, and the visual angle of the teacher's face was 10.6° (see Figure 4.4 (C)). The student was asked to look the teacher in the eye, to listen carefully to the teacher's description, and to learn the content. Only back-channel signals were permitted, and the constraints for the experiment were the same as in the face-to-face communication condition.



[redacted] [redacted] [redacted]

(C)



[redacted]

[redacted]

Figure 4.4. Schematic illustration of the Face-to-face Communication and Remote Communication Conditions. (A) shows the position of a small three-axis acceleration sensor. The sensor (4.5 cm × 4.0 cm) had a sampling frequency of 100 Hz, and it was attached to the forehead of each participant. Head movement was defined as a movement in the vertical and longitudinal direction. (B) shows the experimental situation in the face-to-face communication condition, and (C) shows the experimental situation in the remote communication condition.

4.2.5 Data Analysis

Detection of Phase Difference

Time-series data on the acceleration of head movements in three axes were recorded with a sampling frequency of 100 Hz. Here, we define a head movement as in a previous study, which is a movement in the vertical (superior and inferior) and longitudinal (anterior and posterior) directions (Saiga, et al., 2010). Thus, we only analyzed the two directions of acceleration shown in Figure 4.4 (A). The time-series data of the norm of the accelerations in the vertical and longitudinal directions (x, z) were calculated as

$$a(t_i) = \sqrt{a_x^{\,2}(t_i) + a_z^{\,2}(t_i)} \quad \text{for } i = 0, 1, 2, \ldots. \tag{1}$$

The interval between t_i and t_{i+1} is 10 ms, which is equal to the temporal resolution of the device. As there are differences between individuals in the strength of their nods, $a(t_i)$ was normalized by

$$a'(t_i) = \frac{a(t_i) - \bar{a}}{\sigma_a}. \tag{2}$$

Here, $\bar{a}$ and σ_a are calculated as

$$\bar{a} = \sum_{t_i \in T} \frac{a(t_i)}{|T|}, \tag{3}$$

$$\sigma_a = \sqrt{\frac{\sum_{t_i \in T} \left(\bar{a} - a(t_i)\right)^2}{|T| - 1}}, \tag{4}$$

where T represents the total measurement period in each pair. The time-series data $a'(t_i)$ were smoothed with a moving average of 100 ms to reduce fluctuations due to signal distortion. In a previous study, the durations of posture shifts in head movements were around 400 ms (Hadar, et al., 1984). The moving average of 100 ms means a minimum unit of the same order in the durations of posture shifts in head movements. We calculated the time-series data $a'(t_i)$ as follows

$$\overline{a'}(t_i) = \frac{1}{11}\sum_{l=i}^{i+10} a'(t_l) \quad \text{for } i = 0, 1, 2, \ldots. \tag{5}$$

When head nods occurred, the local maximum values, hereafter called peaks, existed in time-series data $\overline{a'}(t_i)$. We therefore defined the peak acceleration as the $\overline{a'}(t_i)$ that satisfies the following inequality:

$$\overline{a'}(t_i) - \overline{a'}(t_{i\pm1}) > 0. \tag{6}$$

To extract only reliable signals of head nods, we used a threshold amplitude for $\overline{a'}(t_i)$ of 2.0 or more. Peaks of 2.0 or more constituted approximately 6% of the total acceleration peaks in all students' head motions. We used the video data to confirm visually that peaks of 2.0 or more actually corresponded to head nods. Thus, we imposed the following conditions on $\overline{a'}(t_i)$:

$$\overline{a'}(t_i) \geq 2.0. \tag{7}$$

After we detected peaks in the acceleration of head nods by student and teacher, we defined the phase difference as the minimum temporal difference ($t_j - t_i$) from the time (t_i) of a peak in acceleration of the teacher's head nods to that (t_j) of the student. The range of the phase difference was limited to 1.0 s because it has been reported that the maximal delay time for nonverbal synchronization is 1.0 s (Komori and Nagaoka, 2010). Therefore, we imposed the following restriction, in addition to conditions (6) to (7), on the definition of phase difference:

$$-1.0s \leq t_j - t_i \leq 1.0s. \tag{8}$$

In the remote communication condition, although the time-series data on the acceleration of teachers' and students' head movements were measured in real time, there was a delay in the transfer of the data from the video camera to the television. Although the students were unaware of this delay (they perceived the delayed information and reacted to it as if it were in real time), we needed to measure the delay time and to include it in our calculation of phase difference. To measure the delay time, we transmitted video camera images of a software stopwatch (Online Stopwatch, temporal resolution: 1ms) on a computer screen to the television. We took simultaneous pictures of the time depicted on the stopwatch on the computer screen and the one on the television screen and used the difference between them as the delay time. The mean delay time for 50 trials was approximately 160 ± 13 ms (Mean ± SD). Therefore, the acceleration data for the teacher corresponded to the acceleration data for the student with a time delay of 160 ms in data processing.

Analysis of Synchronization

Body movement synchronization was defined as the phase difference distribution during the entire communication period. Therefore, the synchronization characteristics are described using statistical analyses of the phase difference distribution of head nodding over the whole measurement period. Relative frequency is used in the phase difference distribution because the total frequency in the phase difference distribution differed between participant pairs. Specifically, the four statistical measurements are: density, mean phase difference, standard deviation (SD) and kurtosis. First, we introduced the density of the frequency of phase difference, defined as the frequency per minute within each pair. Density is an indicator of synchronization activity (frequency of synchronization). Second, we introduced the mean phase difference, defined as the mean of the distribution. The mean phase difference is an indicator of the synchronization direction (tendency of leader–follower relationship), that is, whether the speaker or the listener leads the body movements in the synchronization built during communication. Third, we introduced the SD, defined as the spread of the phase difference distribution. Fourth, we introduced kurtosis, defined as the degree of convergence to the mean phase difference in the distribution. The SD and kurtosis are indicators of the synchronization strength (stability of the tendency).

4.3 Results

Head nod synchronization is detected for each pair of participants using the phase difference distribution. Figure 4.5 illustrates typical time series data for head nods in the face-to-face communication condition. The relative distribution the phase difference of all student–teacher pairs was plotted because the total frequency in the phase difference distribution differed between participant pairs. Figure 4.6 shows the results for each pair in the face-to-face communication condition (also see Table A–13) and Figure 4.7 shows the total results from the face-to-face communication condition (see also Table A–15). In Figures 4.6 and 4.7, the horizontal axis represents the phase difference when head nods occur, and the vertical axis indicates the relative frequency of head nods. Negative values on the horizontal axis indicate that the student's head nod occurred before the teacher's, whereas positive values indicate the reverse.

Table A-13 shows the results for each pair in the face-to-face communication condition. In the face-to-face communication condition, as shown in Figure 4.7, the phase difference distribution in head nods was symmetric and converged on the vicinity of the center. Specifically, the region of relative frequency where the phase differences take negative values accounted for 31.8%. This means not only that the teacher's movement leads the student's movement, but also that the student's movement leads that of the teacher. In the face-to-face communication condition, the mean density of synchronized head nods across pairs was 9.2 nods/min (SD = 4.2 nods/min). The overall mean (across pairs) of the mean phase differences was 110 ms, and the mean of the standard deviations (SDs) across pairs was 320 ms. The mean kurtosis across pairs was 1.3 (SD = 1.7).

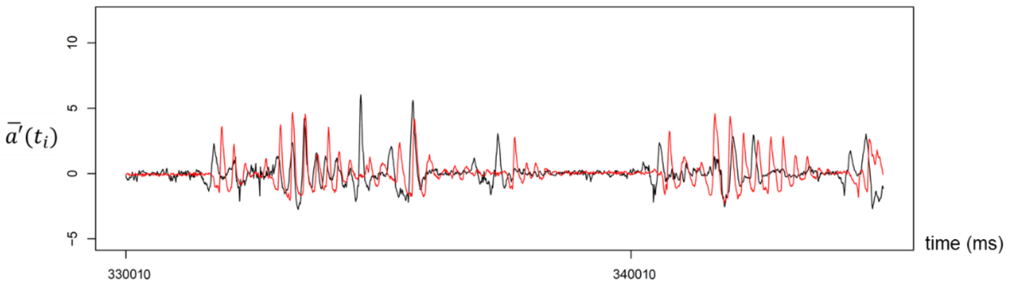


Figure 4.5. Typical Time Series Data for Head Nods in the Face-to-face Communication Condition. The black line indicates the teacher's acceleration data, and the red line shows the student's acceleration data.

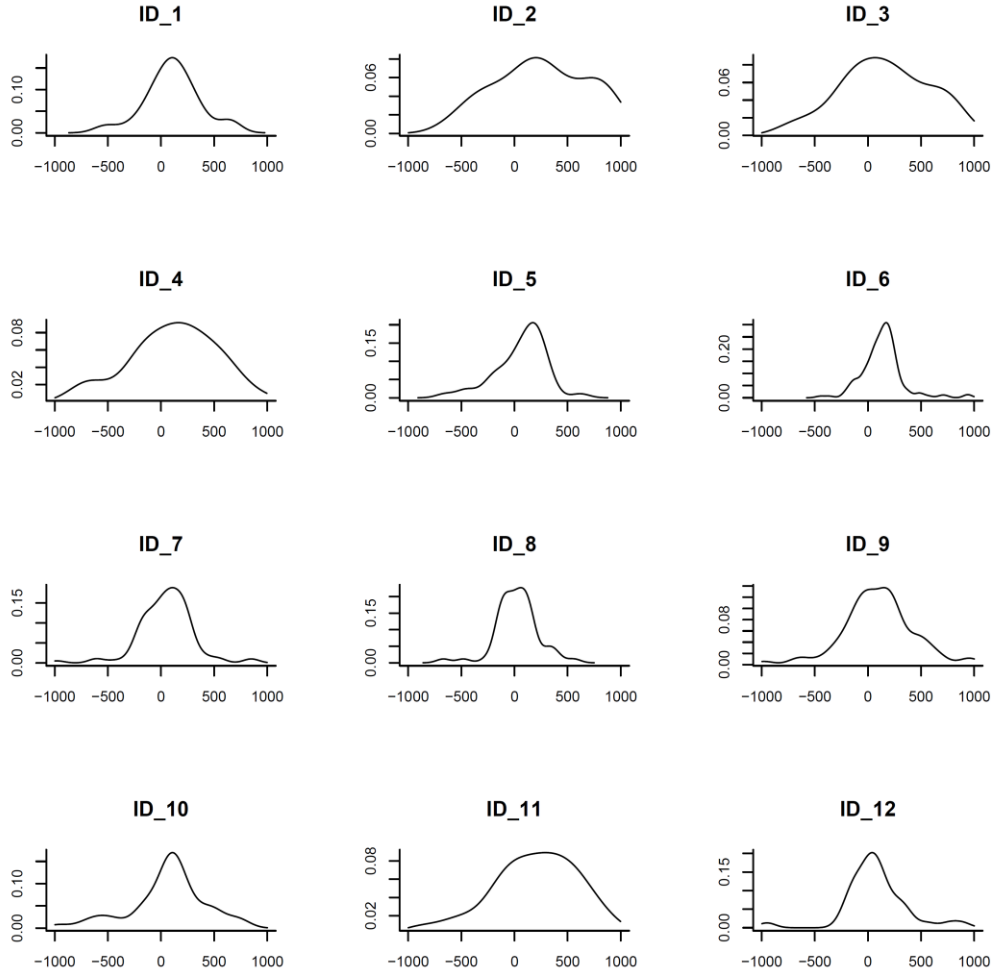


Figure 4.6. Distribution of the relative frequency of synchronized head nods for each pair in the face-to-face communication condition. The horizontal axis represents the phase difference when head nod synchronization occurred, and the vertical axis indicates the relative frequency of head nod synchronization. Negative values on the horizontal axis indicate that the student’s head nod occurred before that of the teacher, whereas positive values indicate the reverse.

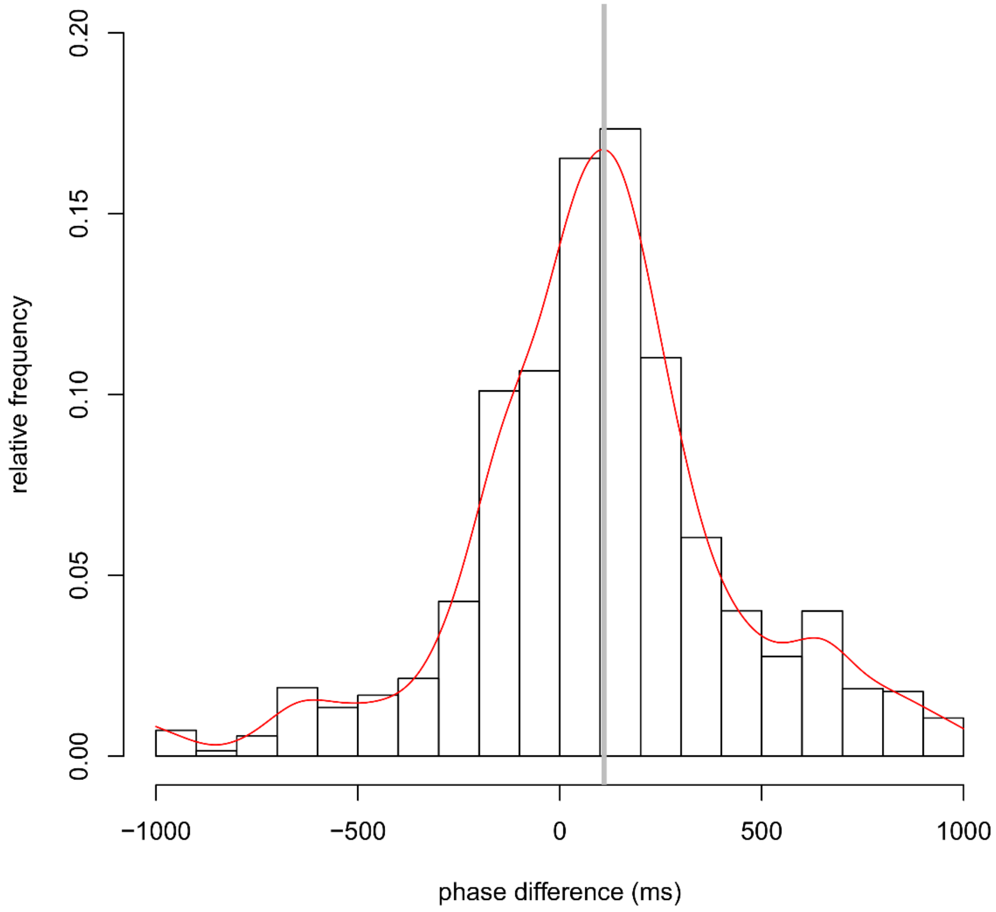


Figure 4.7. Total results from the face-to-face communication condition. Distribution of the mean relative frequency of synchronized head nods across all pairs in the face-to-face communication condition. A smoothing spline curve (red line) is fitted to the mean relative frequency of synchronized head nods across all pairs and the vertical gray line shows the mean phase difference in face-to-face communication. The horizontal axis represents the phase difference when head nod synchronization occurred, and the vertical axis indicates the frequency of head nod synchronization. Negative values on the horizontal axis indicate that the student's head nod occurred before that of the teacher, whereas positive values indicate the reverse.

Figure 4.8 illustrates typical time series data for head nods in the remote communication condition. Figure 4.9 shows the results for each pair in the remote communication condition (also see Table A–14) and Figure 4.10 shows the total results from the remote communication condition (see also Table A–15). In the remote communication condition, the phase difference distribution in head nod synchronization converged on the vicinity of the center. Specifically, the region of relative frequency where the phase differences take negative values accounted for 40.2%. The mean density of synchronized head nods was 4.6 nods/min (SD = 1.7 nods/min). The overall mean (across pairs) of the mean phase differences was 80 ms, and the mean of the SDs across pairs was 430 ms. The mean kurtosis across pairs was 0.1 (SD = 0.9).

Unpaired t-tests indicated that the density of synchronized head nods in the face-to-face communication condition was significantly higher than that in the remote communication condition (t(22) = 3.420, P = 0.002; see Figure 4.11(A)), but the mean phase differences in synchronized head nods did not show a significant difference between the conditions (t(22) = 0.937, P = 0.359, see Figure 4.11(B)). Unpaired t-tests also revealed that SDs in the face-to-face communication condition were significantly smaller than those in the remote communication condition (t(22) = –3.405, P = 0.003; see Figure 4.11(C)) and kurtoses in the face-to-face communication condition were significantly higher than those in remote communication condition (t(22) = 2.098, P = 0.048; see Figure 4.11(D)).

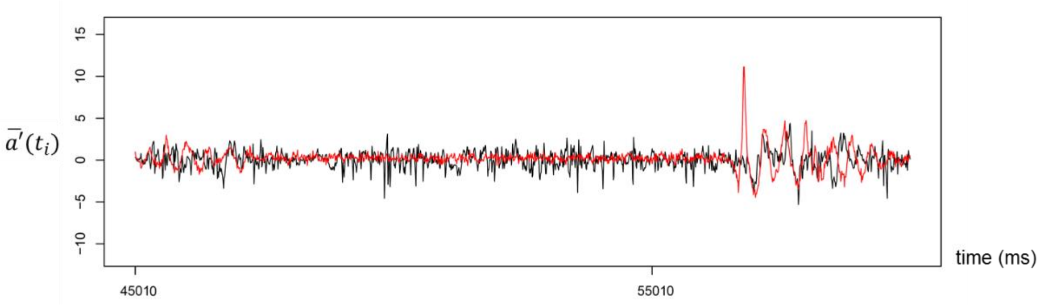


Figure 4.8. Typical Time Series Data on Head Nods in the Remote Communication Condition. The black line indicates the teacher’s acceleration data, and the red line shows the student’s acceleration data.

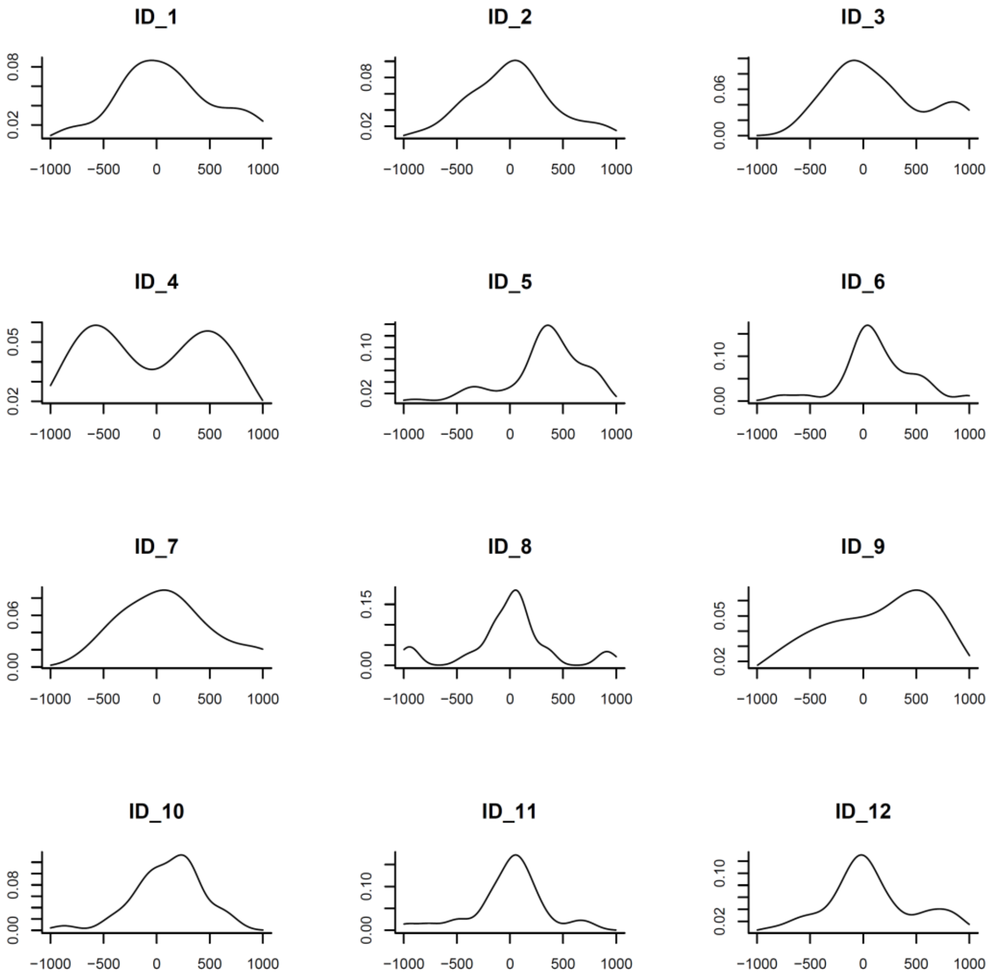


Figure 4.9. Distribution of the relative frequency of synchronized head nods for each pair in the remote communication condition. The horizontal axis represents the phase difference when head nod synchronization occurred, and the vertical axis indicates the relative frequency of head nod synchronization. Negative values on the horizontal axis indicate that the student’s head nod occurred before that of the teacher, whereas positive values indicate the reverse.

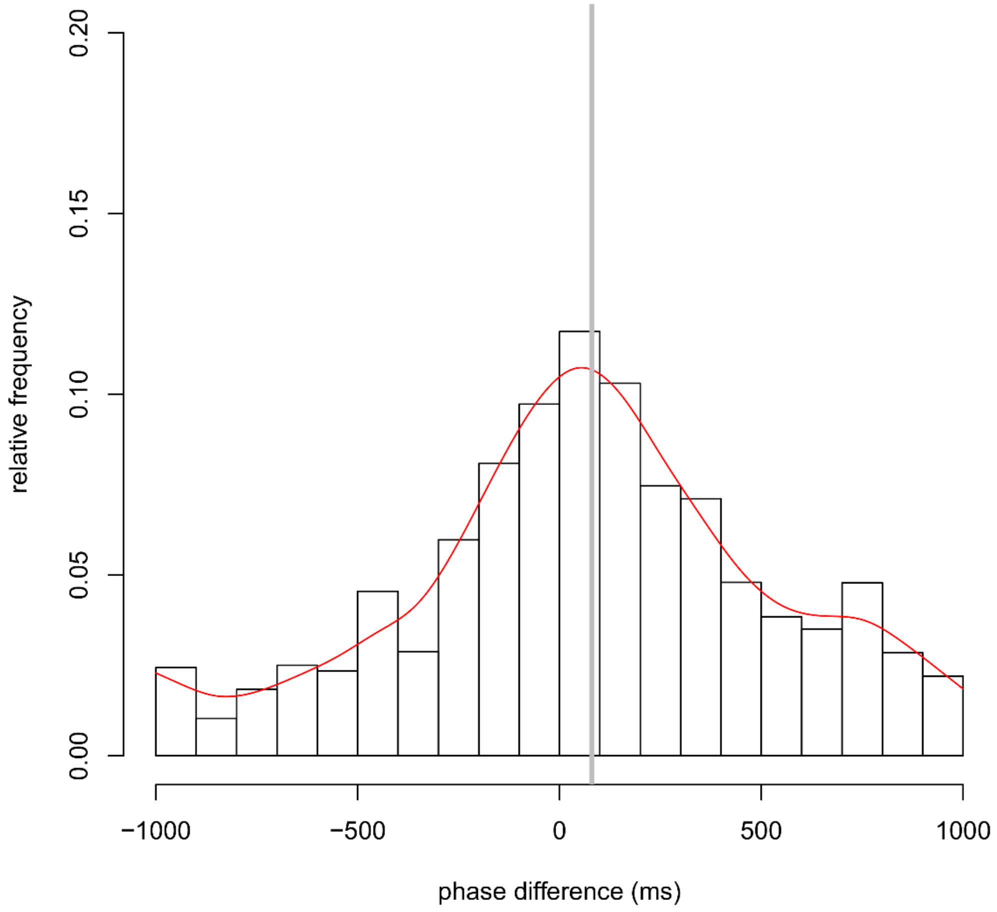


Figure 4.10. Total results from the remote communication condition. Distribution of the mean relative frequency of synchronized head nods across all pairs in the remote communication condition. A smoothing spline curve (red line) is fitted to the mean relative frequency of synchronized head nods across all pairs and the vertical gray line shows the mean phase difference in remote communication. The horizontal axis represents the phase difference when head nod synchronization occurs, and the vertical axis indicates the frequency of head nod synchronization. Negative values on the horizontal axis indicate that the student's head nod precedes that of the teacher, whereas positive values indicate the reverse.

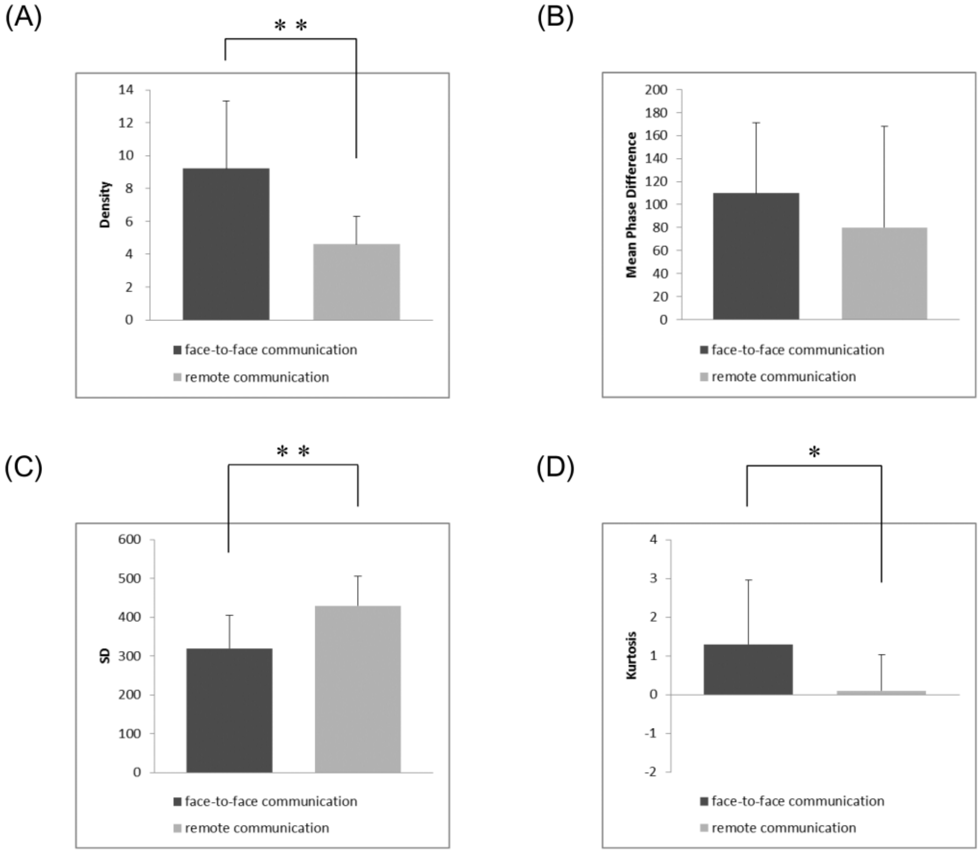


Figure 4.11. Comparison of results between the Face-to-face Communication and Remote Communication Conditions. (A) shows the mean density of synchronized head nods in the face-to-face communication and remote communication conditions. The error bars represent the standard deviations of the means. (B) shows the mean phase difference in the face-to-face communication and remote communication conditions. The error bars represent the standard deviations of the means. (C) shows the mean SDs in the distribution of the phase differences in the face-to-face communication and remote communication conditions. The error bars represent the standard deviations of the means. (D) shows the mean kurtosis in the distribution of the phase differences in the face-to-face communication and remote communication conditions. The error bars represent the standard deviations of the means, *: p < .05, **: p < .01, unpaired t-test.

4.4 Discussion

This study is focused on the influence of visual motion on simultaneity perception using body movement synchronization. To confirm the influence of visual motion, Body movement synchronization is detected in direct face-to-face and remote communication settings using a new definition of phase difference. In this study, the phase difference of head nods is defined during the entire communication period and the characteristics of head nod synchronization is defined as the phase difference distribution. Specifically, the phase difference distribution was characterized using four statistical measurements: the density, the mean phase difference, the SD, and the kurtosis. To verify the validity of the influence of visual interaction and our definition for detecting synchronization, we applied our method to two human communication situations: the face-to-face communication condition with visual interaction and the remote communication condition with unidirectional visual perception. As a result, the difference between the phase difference distributions in these communication situations was clearly shown. Although the mean phase differences in head nods did not differ significantly between the face-to-face communication and remote communication conditions, there were significant differences in the densities, the SDs and kurtoses in the phase difference distributions of head nod synchronization between the face-to-face communication and remote communication conditions. These results can be used to clarify the influence of visual motion on simultaneity perception in human communication. In addition, these results can used to clarify the effectiveness of phase difference as a measure of body movement synchronization and the characteristics of body movement synchronization through the features of the phase difference distribution.

4.4.1 The Influence of Visual Motion on Simultaneity Perception in Human Communication

Our findings clarified some of the differences in head nod synchronization between face-to-face communication with visual interaction and remote communication without visual interaction. Specifically, the density, the SD and the kurtosis of the phase difference distribution differed between the face-to-face communication condition with visual interaction and the remote communication condition with unidirectional visual perception. Thus, visual interaction in the face-to-face communication condition led to a large density of synchronized head nods and a small spread (small SD and large kurtosis) compared with the remote communication condition. This means that visual interaction resulted in higher synchronization activity and strength. Most importantly, these differences clearly showed the mechanism of body movement synchronization in human communication. Schmidt et al. (1990) reported that visually coupled perceptions between individuals is important for the coordination of body movements. In both communication conditions in the present study, the teachers presented the same auditory information to the students. However, these conditions differed in their visual modality, because the teachers could see the students' back-channel signals in the face-to-face communication condition, but this information was not available to the teachers in the remote communication condition (see Figure 4.12). This interaction through the visual channel may contribute to mutual entrainment in nonverbal synchronization, because synchronization phenomena are established by the mutual entrainment mechanism based on interaction between nonlinear oscillators from a theoretical viewpoint (Kuramoto, 1984; Acebron, et al., 2005). This finding will play an important role in the elucidation of the mechanism of nonverbal synchronization in face-to-face communication and the application of remote communication technologies.

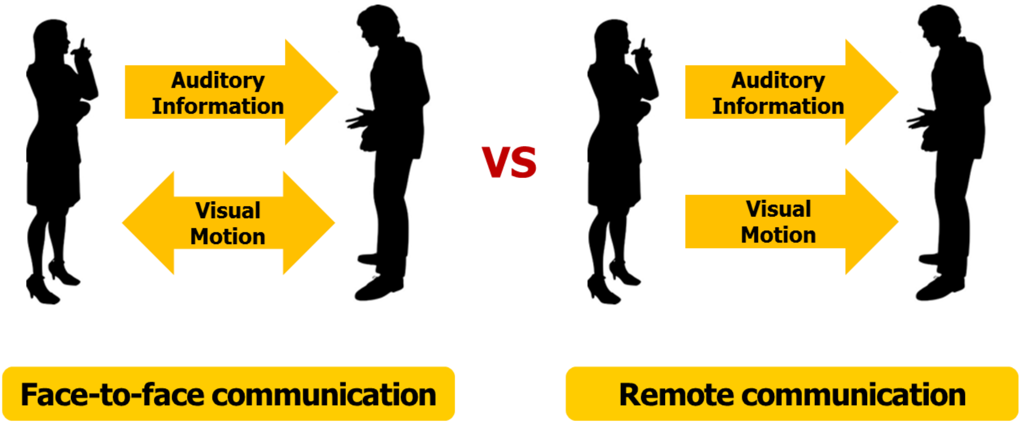


Figure 4.12. The differences in body movement synchronization between direct face-to-face and remote communications. In both of the experiments in the present study, the teachers presented the same auditory information to the students. However, the two experiments differed in the visual modality, because the teachers could see the students' back-channel signals in the direct face-to-face communication experiment, but this information was not available to the teachers in the remote communication experiment.

However, there was no difference in the mean phase difference between face-to-face communication with visual interaction and remote communication with unidirectional visual perception. This means that visual interaction in the head nod synchronization did not affect the mean phase difference. The mean phase difference is an indicator of the synchronization direction, that is, whether the speaker or listener leads the body movements in the synchronization built during communication. The speaker's head nods tended to slightly lead the listener's head nods in both communication conditions. In recent years, the mechanism of nodding in face-to-face communication has been reported. Bavelas et al. (2002) reported that the nodding of the listener occurred in a gaze window, which is a temporal window of mutual gaze created by the speaker looking towards the listener. In addition, according to Stivers (2008), nods by a listener act as a sign of alignment with the activity of speaking and affiliation through a claim of access to the speaker's stance, either indirectly or directly. These studies well represent the mechanism of the occurrence and function of nodding. However, in this study, the synchronization characteristic of head nods was detected even in the remote communication condition without mutual gaze related to visual interaction, in which there was no difference in the mean phase differences between the face-to-face and remote communication conditions. This therefore shows the synchronization direction of head nods may be attributed to the listener's alignment, that is, the listener's adaptive behavior to the speaker's multimodal behavior, even in remote communication.

4.4.2 Automatic Processing in Body Movement Synchronization

Further, in both communication conditions, the phase difference distribution in head nod synchronization was symmetric and converged on the vicinity of the center. Specifically, a region of 30–40% in relative frequency corresponded to the region in which the student’s head nods led those of the teacher. This means that the listener’s movement is not always aligned with the speaker’s movement. In particular, it is known that when a simple stimulus such as a letter or symbol is presented, the reaction time of body movements is within a range of 380 to 420 ms (Laming, 1968; Luce, 1986; Sternberg, 1969). However, in this study, the mean phase difference was 110 ms with an SD of 320 ms in the face-to-face communication condition. This shows a narrower temporal feature in nonverbal synchronization compared with the reaction time of body movements in the previous studies. This suggests that the temporal feature of head nod synchronization may be due to automatic processing of body movements in human communication.

In recent years, mutual gaze in face-to-face communication has been reported as an important factor for the mechanism of nodding. Bavelas et al. (2002) reported that the nodding of the listener occurred in a gaze window, which is a temporal window of mutual gaze created by the speaker looking towards the listener. In addition, according to Stivers (2008), nods by a listener act as a sign of alignment with the activity of speaking and affiliation through a claim of access to the speaker’s stance, either indirectly or directly. These studies well represent the mechanism of the occurrence and function of nodding. However, in this study, the synchronization of head nods was detected even in the remote communication condition without mutual gaze. This shows that subconscious and automatic processing of nonverbal communication may be attributable to the synchronization of head nods, even in remote communication.

In the present study, we focused on the subconscious and automatic aspect of head nods in human communication. However, the mechanism of head nods has a conscious, reactive and semantic aspects (Bavelas et al., 2002; Stivers, 2008). Therefore, in the future we need to examine the influence of the conscious, reactive and semantic aspects. According to this study, the mean phase difference, SD and kurtosis can indicate differences in the phase difference distribution between the conscious and reactive head nods and the synchronization of subconscious and automatic head nods. Then the mean phase difference, SD and kurtosis in the subconscious and automatic head nods in synchronization may become smaller than that for conscious and reactive head nods. In addition, it is expected that the phase difference distribution for subconscious and automatic head nods in synchronization will lead to a smaller spread with a small SD and large kurtosis compared with the distribution of those of the conscious head nods.

4.4.3 Newly-devised Detection Method

With respect to a newly-devised method of detecting nonverbal synchronization, our findings showed a narrower temporal window for nonverbal synchronization compared with a previous study (Komori and Nagaoka, 2010). Simple reaction time, which participants press a button when light or sound is presented, was approximately 160 ms for an auditory stimulus and 190 ms for a visual stimulus (Kosinski, 2013). However, when a stimulus was complicated such as a letter or symbol, the reaction time increased within a range of 380 to 420 ms (Laming, 1968; Luce, 1986; Sternberg, 1969; Kosinski, 2013). With regard to head nods, Watanabe (1985) reported that the phase difference in head nod synchronization was 700 ms to 1400 ms, when the voice of the speaker affects the nodding of the listener. In addition, the phase difference in synchronization was approximately 800 ms to 2.0 seconds when the nodding of the listener affects the voice of the speaker. With regard to nonverbal synchronization, Komori (2010) reported that the time lag between the amplitude of the body movements of a counselor and a client occurred within a range of ± 1000 ms. In particular, there was a tendency for the counselors' body movements to be delayed by 500 ms compared with those of clients. However, in this study, the mean phase difference was 110 ms, and the SD was 320 ms. In addition, in face-to-face communication condition, a region with a high occurrence frequency of head nod synchronization was that within a range of 100 ms to 200 ms and the head nod synchronization of approximately 70% showed in a range of - 200 ms to 300 ms. Therefore, we confirmed that the high occurrence frequency of head nod synchronization showed in a range of - 200 ms to 300 ms. Moreover, the overall mean (across pairs) of the mean phase differences was 110 ms and the mean of the standard deviations (SDs) across pairs was 320 ms. Both values are within the range of high occurrence frequency of head nod synchronization. These results are obtained with a newly-devised method of detecting nonverbal synchronization based on a measure of phase difference with a high degree of temporal resolution, which was defined as the phase difference in the peak amplitudes between rhythmic body movements. Thus, this method can detect the accurate temporal properties of nonverbal synchronization.

4.4.4 New Approach for Defining Synchronization

As discussed above, phase difference is used to detect body movement synchronization, and the density, mean phase difference, SD and kurtosis in the phase difference distribution are used to evaluate the degree of body movement synchronization. Therefore, there is a need to further define indicators to characterize nonverbal synchronization properly. In the field of cognitive psychology, it has been found that there exists a point of subjective simultaneity (PSS), which is the maximal simultaneity in multisensory processing by a human perceptual system. The PSS is commonly used, which is an indicator of subjective simultaneity in multisensory processing by a human perceptual system. In particular, it has been reported that the PSS differs from physical simultaneity in multisensory integration (Vroomen and Keetels, 2010; Spence et al., 2001; Spence et al., 2010; Kwon, 2014). Therefore, they have claimed the importance of the relationship between physical simultaneity and subjective simultaneity according to type of stimulation. The PSS is obtained by the mean of a distribution obtained by calculating the percentage of "simultaneous" responses to various stimulus onset asynchronies (SOAs) between stimuli. For example, in audiovisual simultaneity, the PSS was usually shifted toward a visual-lead stimulus (Vroomen and Keetels, 2010; Spence et al., 2001; Spence and Parise, 2010; Kwon et al., 2014; Jakowski et al., 1990; Kanabus et al., 2002; Zampini et al., 2003; Zampini et al., 2005). This indicates that the visual stimulus should be presented faster than the auditory stimulus for perceptual audiovisual simultaneity. Interestingly, this study reveals that same tendency showed between the phase difference in nonverbal synchronization and the PSS in multisensory integration. That is, the mean phase difference in nonverbal synchronization of head nods was not zero (i.e., physically perfect synchronization), and is shifted toward the head nods of the listener lagging behind those of the speaker. This means that nonverbal synchronization of human communication differs from physically perfect synchronization. Therefore, the mean phase difference can be called “the point of subjective synchronization”, and it can be used as an indicator of subjective synchronization in human communication.

Moreover, there is another indicator in addition to the PSS in multisensory integration. This is called the “temporal window of integration,” which is the range of SOAs within which we perceive two pieces of sensory information as the same event (Spence et al., 2001; Spence and Parise, 2010; Kwon et al., 2014; Zampini et al., 2005; Morein-Zamir et al., 2005; Keetels and Vroomen, 2005; Wassenhove et al., 2007). The temporal window of integration is calculated as the standard deviation (SD) of a distribution derived by calculating the percentage of “simultaneous” responses to various SOAs between stimuli. It is known that when using a simple pair of a sound stimulus and a flash stimulus, the temporal window of integration showed within a small range between approximately 30 ms and 110 ms (Zampini, et al., 2003; Zampini, et al., 2005; Morein-Zamir, et al., 2003; Keetels and Vroomen, 2005). On the other hand, it has been reported that the temporal window of integration in audiovisual speech is approximately 200 ms (Wassenhove, et al., 2007). From this perspective, the SD of the spread of the phase difference distribution in this study can be regarded as an indicator of “the temporal window of synchronization” in human communication. In this study, the distribution of nonverbal synchronization in face-to-face communication differed from that in remote communication. The spread in the phase difference distribution in face-to-face communication is smaller than that in remote communication, and the convergence in the phase difference distribution in face-to-face communication was greater than that in remote communication. However, this definition is an open question. Figures 4.7 and 4.10 show that the phase difference distribution changes in shape for every 100 ms (also see Figures 4.13 and 4.14). Thus, it can be regarded as a temporal unit of the temporal window of synchronization. In future, we need to examine whether the SD of the phase difference distribution should be expressed in this unit in the temporal window of synchronization.

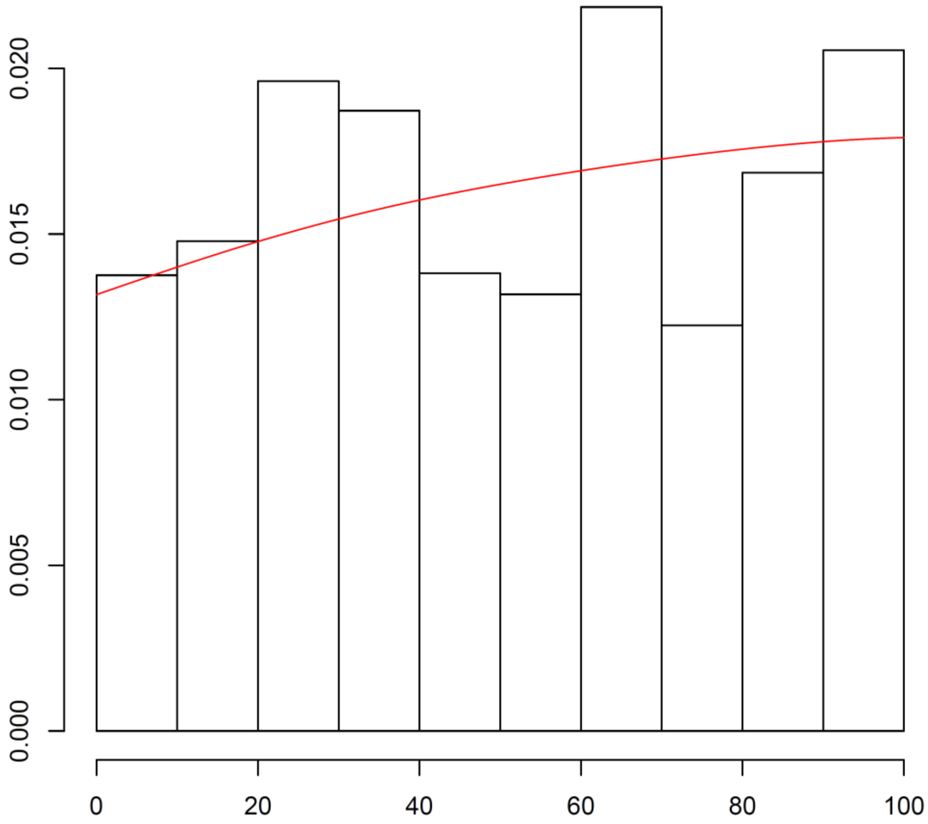


Figure 4.13. Distribution of the mean relative frequency of synchronized head nods within a range of 0 to 100 ms across all pairs in face-to-face communication. Red lines show smoothing spline curves.

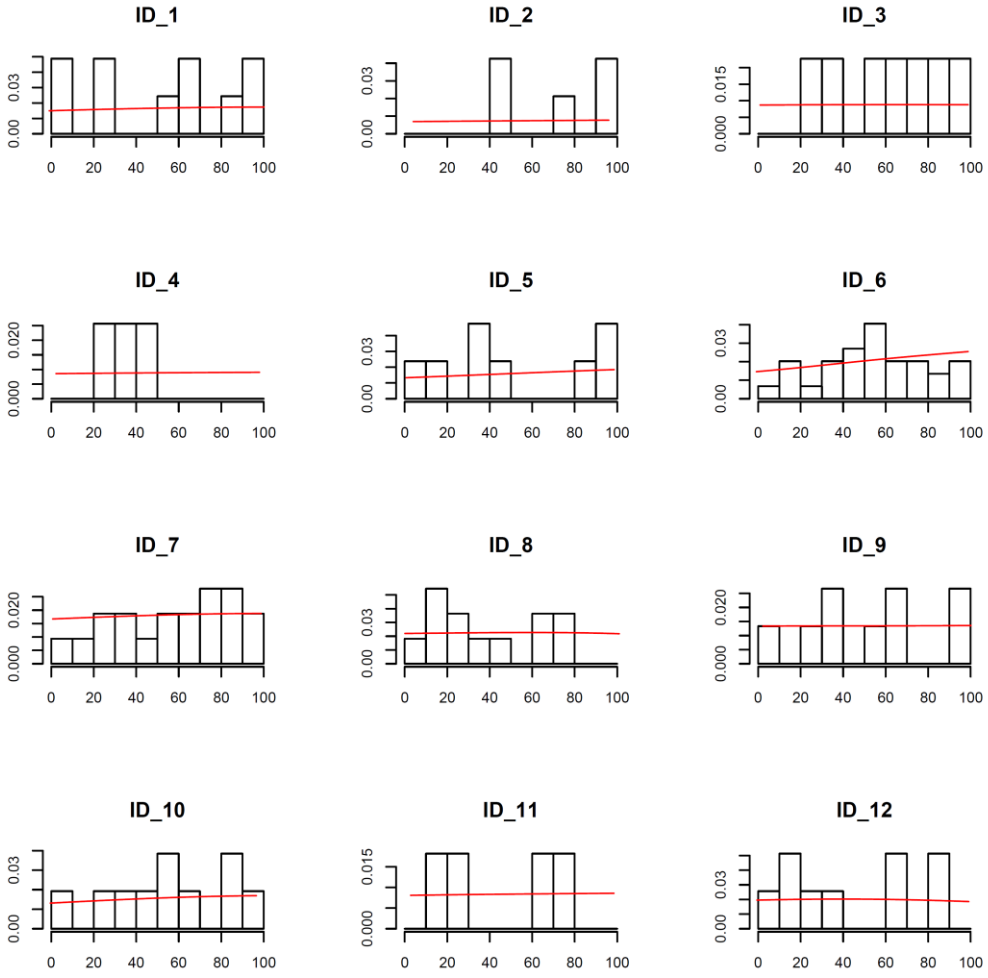


Figure 4.14. Distribution of the relative frequency of synchronized head nods within a range of 0 to 100 ms for each pair in face-to-face communication. Red lines show smoothing spline curves.

The point of subjective synchronization in nonverbal communication was shifted to a greater extent than that in multisensory integration (see Table 4-1). In addition, as mentioned above, the temporal window of integration in audiovisual speech was larger than that in an experiment using a simple pair of sound and flash stimuli (Zampini, et al., 2005; Wassenhove, et al., 2007), (see Table 4-1). This means that conversation attributes a factor that contribute to the larger range of the audiovisual temporal integration. In this study, we investigated nonverbal synchronization in audiovisual speech with body movements. We found that the point of subjective synchronization was approximately 107 ms and the temporal window of synchronization was approximately 301 ms. Consequently, the point of subjective synchronization in nonverbal communication and the temporal window of synchronization are larger than those of multisensory integration in audiovisual speech. This may be attributable to the effect of body movements as an additional factor in multisensory integration in audiovisual speech.

Table 4-1. Changes in PSS and SD in multisensory integration in previous studies and Nonverbal synchronization.

	PSS	SD
Simple stimuli	19 ms	114 ms
Audiovisual speech	23 ms	205 ms
	Mean phase difference	SD
Nonverbal synchronization	107 ms	310 ms

- Simple stimuli: audiovisual simultaneity judgments in simple pair of a sound stimulus and a flash stimulus.
- Point of subjective simultaneity (PSS) means a maximal simultaneity that a pair of audiovisual stimuli were perceived as simultaneous, and SD indicates temporal window of multisensory integration.
- Audiovisual speech: audiovisual simultaneity judgments between auditory speech stimuli and visual speech stimuli.
- Nonverbal synchronization: the synchronization of head nods in face-to-face communication. Mean phase difference refers to mean phase difference in peaks of acceleration of head nods in face-to-face communication, and SD is the standard deviation in the phase difference distribution in peaks of head nods in face-to-face communication.

4.4.5 Perspectives on Remote Communication Technologies

From an experimental viewpoint, two factors lead to such differences between direct face-to-face and remote communications. Technologies developed in recent years enable us to communicate remotely as well as in person (Harboe, et al., 2008; Cesar, et al., 2008). Although remote communication has been developed to approximate face-to-face communication, remote communication, especially via television, remains inadequate compared with face-to-face communication (van der Kleij, et al., 2005; van der Kleij, et al., 2009). The first factor is the spatial separation in remote communication (Morris, 2004; Jones, 1993). Although face-to-face communication between partners occurs in the same room, remote communication is carried out in different rooms. This shared space in the former condition may contribute to the differences in body movement synchronization between direct face-to-face and remote communications. The second factor is the lack of authenticity in remote communications (van der Kleij, et al., 2005; van der Kleij, et al., 2009). Remote communications have been developed to approximate face-to-face communication, but the reality created by such interactions still falls short of face-to-face communication. Therefore, the lack of authenticity in remote communications may have an effect on body movement synchronization. These factors should be studied to elucidate the mechanisms of face-to-face and remote communications.

4.4.5 Future Research

In this study, we applied our definition to human communication in which the roles of speaker and listener were defined. We analyzed the phase difference distribution between the face-to-face communication condition with visual interaction and the remote communication condition without visual interaction, to clarify the influence of visual interaction in the synchronization of head nods between the speaker and the listener. However, in the future it will be necessary to examine other factors such as mutual talk to clarify the influence of other interactions as a cause of synchronization in human communication. In addition, there is a need to examine the verbal factor in which the listener can only hear the speaker but has no visual access to the speaker in order to determine the unique influence of verbal and nonverbal behavior. Also, data in this study were obtained only from Japanese conversations and head nods. Therefore, there is a further need to examine the influence of other languages, different cultures and nonverbal signals. Furthermore, our data were obtained only from Japanese conversations and head nods. Therefore, there is a further need to examine the influence of other languages, different cultures and nonverbal signals. In this study, the threshold for head nods was an amplitude of $a'(t_i)$ of 2.0 or more. However, there may have been weak head nods in the acceleration peaks of 2.0 or less. Therefore, it should be analyzed this possibility in future work. Our findings will prompt research on future communication technology based on nonverbal synchronization in face-to-face and remote communication. We believe that these findings are useful in detecting nonverbal synchronization in various human communication situations.

Chapter 5

General Discussion

5.1 Main Findings

This dissertation addresses the influence of visual motion information on simultaneity perception. This dissertation is focused on visual apparent motion with respect to motion information and TOJ task with regard to simultaneity perception. In addition, body movement synchronization is investigated to confirm the characteristics of simultaneity perception in human communication. The main aims of this dissertation were (1) to examine how visual apparent motion affects audiovisual simultaneity perception, (2) to investigate what mechanisms contribute to the finding that apparent motion affects simultaneity perception, (3) to confirm the influence of visual motion on simultaneity perception in human communication.

In Chapter 2, the purpose was to investigate how visual apparent motion affects audiovisual temporal perception. Two types of TOJ tasks was examined to confirm whether visual apparent motion has an effect on an audiovisual TOJ task. Participants conducted audiovisual TOJ tasks in the apparent motion condition with two flashes, and in the normal condition with a single flash, which is the conventional condition of a TOJ task. The results of Chapter 2 show that the PSS in the apparent motion condition was shifted toward a sound-lead stimulus, which is closer to physical simultaneity and the JND in the apparent motion condition was smaller than that in the normal condition. This means that visual apparent motion contributes to very precise perceptions of temporal simultaneity, which is closer to physical simultaneity, in audiovisual integration, and greater temporal resolution than that which occurs in the normal condition and the previous studies.

In Chapter 3, the goal was to investigate what mechanisms contribute to the finding of Chapter 2 that visual apparent motion affects audiovisual simultaneity perception. Three possible mechanisms were considered in the Chapter 3 and they were examined in three sections, respectively.

In Section 1, the purpose was to examine the influence of amount of visual stimulation by eliminating the influence of the amount of flash stimulation, because it remained unclear whether the Chapter 2’s results were influenced by differences in the amount of visual stimulation between the apparent motion condition with two flashes and the normal condition with a single flash. As a result, the PSS and JND obtained in the apparent motion condition differed from those obtained in the successive condition as non-motion perception, which included the same amount of visual stimulation in the apparent motion condition. Similar results were obtained in the apparent motion conditions between Chapter 2 and Section 1 in Chapter 3. Also, the results obtained in the successive condition were similar to those obtained in the normal condition in Chapter 2. This means that the amount of visual stimulation made no difference in the apparent motion condition and the PSSs and JNDs in audiovisual temporal perception differed according to whether visual apparent motion was present or absent.

In Section 2, the purpose was to investigate the influence of prediction as a higher-order brain function, because apparent motion was produced by a constant interval between two flashes. Therefore, it was necessary to eliminate the influence of prediction by randomizing the intervals between the two visual stimuli. The results of Section 2, which eliminated the effect of prediction, were no different from those obtained in the apparent motion condition in Chapter 2. This means that apparent motion was equivalently processed regardless of prediction. In particular, temporal processing during apparent motion perception may result from stimulus-driven effects rather than expectancy-driven effects. It raise the possibility that despite the stimulus-driven effects with the same type of sensory stimuli, the difference between motion and non-motion perceptions influenced temporal order perception and the window of temporal integration in audiovisual processing.

In Section 3, the purpose was to investigate whether motion binding property in visual perception influences audiovisual simultaneity perception. Although two discrete stimuli are presented at appropriate spatiotemporal intervals, we can perceive continuous motion as one moving object. This phenomenon suggests that visual motion perception mechanisms have a binding property that stimuli are perceived as a moving object with spatiotemporal continuity, and a single bounded object is perceived only at certain intervals. Apparent motion is a most suitable phenomenon to show motion binding property in human perceptual system. As a result, the PSS shifted toward a sound-lead stimulus and especially became closer to physical simultaneity (i.e., zero) in the motion binding condition, and the JND in the motion binding condition was smaller than those in the non-motion binding condition. This shows that motion binding property shifted simultaneity toward physical simultaneity, and reduces the temporal window in audiovisual integration. Therefore, it revealed that motion binding property contributes to the accurate perceptions of temporal events in audiovisual simultaneity perception.

In Chapter 4, the purpose was to investigate the influence of visual motion on simultaneity perception in human communication. In addition, Chapter 4 provides a definition of phase difference for the detection of body movement synchronization in human communication. To confirm the influence of visual motion, we detected body movement synchronization in direct face-to-face (bilateral perception of visual motion) and remote communication (unilateral perception of visual motion) settings using a new definition of phase difference. As a result, although the mean phase difference in synchronized head nods did not differ significantly between face-to-face and remote communications, there were significant differences in the density, the SDs and kurtoses in the phase difference distribution of head nod synchronization between face-to-face and remote communications. The findings clarified the differences in body movement synchronization between direct face-to-face and remote communications. In particular, motion feedback loop in the visual channel play an important role in the synchronization mechanism in face-to-face communication. Also, these results are obtained with a newly-devised method of detecting nonverbal synchronization based on a measure of phase difference with a high degree of temporal resolution. Therefore, this study shows the effectiveness of phase difference as a measure of body movement synchronization and it can be used to evaluate body movement synchronization.

5.2 Motion Perception and Simultaneity Perception

Motion perception is a fundamental tool for interaction with a dynamic environment in real time and temporal information of motion is very important for smooth interaction with the dynamic environment. Also, simultaneity perception enables to perceive multiple sensory information as a single event and interact with the environment at the same time. Therefore, the relationship between motion perception and simultaneity perception is important for interaction with the dynamic environment because real time interaction with the dynamic environment is generated by the relationship between motion perception and simultaneity perception.

Chapter 2 is focused on the influence of visual motion on the multisensory temporal integration from an environment. In a dynamic environment, visual motion information is an influential factor in temporal perception and especially is important in multisensory temporal perception. However, it is unclear how visual motion information influences the temporal integration of multisensory perceptions. In Chapter 2, the PSS in the apparent motion condition was shifted toward a sound-lead stimulus, which differs from the PSS in the normal condition. Moreover, the JND in the apparent motion condition was smaller than that in the normal condition. This finding shows visual apparent motion changes temporal simultaneity perception and improves temporal discrimination in audiovisual processing.

Previous studies have reported that the temporal integration of multisensory information depends on the combination of sensory information and they can be changed according to a variety of factors such as stimulus intensity, prediction and attention (Allik and Kreegipuu, 1998; Jakowśki and Verleger, 2000; Mattes and

Ulrich, 1998; Schneider and Bavelier, 2003). In particular, many studies have shown that audition dominates vision on multisensory processing in the time dimension (Gebhard and Mowbray, 1959; Fendrich and Corballis, 2001; Bertelson and Aschersleben, 2003; Morein-Zamir, et al., 2003; Freeman and Driver, 2008; Recanzone, 2003; Vroomen and de Gelder, 2004). In addition, temporal ventriloquism effect has shown that temporal perception of visual stimuli is pulled into that of auditory stimuli (Fendrich and Corballis, 2001; Morein-Zamir, et al., 2003; Scheier, et al, 1999; Vroomen and Keetels, 2006). However, visual apparent motion changes temporal simultaneity perception and improves temporal discrimination in audiovisual processing. Therefore, visual motion has a significant impact on the audiovisual simultaneity perception.

In previous studies, audiovisual information by the same spatial position or congruent audiovisual speech stimuli resulted in that the PSSs were closer to physical simultaneity but the JNDs become larger in audiovisual processing (Wassenhove, et al., 2007; Zampini, et al., 2003). On the other hand, audiovisual information by the spatial separation or incongruent audiovisual speech stimuli resulted in that the PSSs become more distant from physical simultaneity but the JNDs become smaller with higher temporal resolution in audiovisual processing. In other words, audiovisual information by the same spatial position is related to larger temporal window in audiovisual integration and audiovisual information by the different spatial position is engaged in the reduction of the temporal window in audiovisual integration with higher temporal resolution and temporal sensitivity. However, visual apparent motion affects audiovisual simultaneity by shifting simultaneity toward physical simultaneity, and reduces the temporal window in audiovisual integration. This suggests that visual motion information contributes to accurate perceptions of temporal events in the physical world.

5.3 Motion and Amount of Stimulus on Simultaneity Perception

The purpose of Chapter 2 was to investigate whether visual apparent motion affects audiovisual simultaneity perception. As a result, the findings reveal that visual apparent motion affects audiovisual simultaneity by shifting simultaneity toward physical simultaneity, and reduces the temporal window in audiovisual integration. However, it is not clear that what mechanisms contribute to the finding that apparent motion affects simultaneity perception. Therefore, there was a need to confirm the three possible mechanisms, which are 1) influence of amount of stimulus (Section 1 in Chapter 3), the influence of prediction (Section 2 in Chapter 3) and the influence of motion binding property on audiovisual simultaneity (Section 3 in Chapter 3).

In Section 1, the results obtained in the apparent motion condition differed from those obtained in the successive condition, which included the same amount of visual stimulation in the apparent motion condition. Similar results were obtained in the apparent motion conditions between Chapter 2 and Section 1 in Chapter 3. Also, the results obtained in the successive condition were similar to those obtained in the normal condition in Chapter 2. This means that the PSSs and JNDs in audiovisual temporal perception differed according to whether visual apparent motion was present or absent. In Chapter 2, it remained unclear whether the results were influenced by differences in the amount of visual stimulation between the apparent motion condition with two flashes and the normal condition with a single flash. There was a possibility that visual apparent motion influenced perceptions of audiovisual simultaneity, because visual energy increases in the apparent motion condition. In addition, this may change the relative perception of audiovisual simultaneity compared with the use of a single visual stimulus, because double visual stimuli in the apparent motion condition would prime the visual processing system (Spence, et al., 2001; Posner and Nissen, 1976). It is particularly notable that Spence et al. (2001) reported that the temporal processing of modalities was affected not only by the prediction of modality expectancies but also by the quantity of modality-induced stimuli. In Section 1, therefore, the amount of visual stimulation in the two conditions was equalized, and the participants then conducted TOJ tasks in the apparent motion condition and in the successive condition. However, the amount of visual stimulation made no difference in the apparent motion condition.

5.4 Motion and Top-down Factors on Simultaneity Perception

The results of Section 2 in Chapter 3, which eliminated the effect of prediction, were no different from those obtained in the apparent motion condition in Chapter 2. This means that apparent motion was equivalently processed regardless of prediction. With respect to visual prediction and attention, it should be noted that when participants know the specific time at which targets appear, specific attention can be allocated (Coull and Nobre, 1998). Furthermore, it is known that predictable and anticipated information improves temporal resolution and temporal sensitivity (Petrini, et al., 2009). The attention modulates neural activity (O'Craven, et al., 1997; Treue and Martinez Trujillo, 1999), and a faster time course is allocated for motion processing (Busse, et al., 2008). However, the result of unpredictable apparent motion did not differ from that for predictable apparent motion. Therefore, prediction and intention as top-down factors have no effect on the results of Chapter 2.

Many researchers have reported that prediction influences the temporal processing of modalities because temporal processing of expected modalities is faster than that of unexpected modalities (Posner, et al., 1980; Duncan, 1980; Spence and Driver, 1997; Klein, 1977). However, Spence et al. reported that reaction times for a sensory stimulus followed by another sensory stimulus of the same type were faster than when a cross-modal stimulus was expected (Spence, et al., 2001). Therefore, Spence et al. suggested that stimulus-driven and expectancy-driven effects must be distinguished in studies of the temporal processing of sensory modalities (Spence, et al., 2001). The results of Chapter 2 and Section 2 in Chapter 3 show no influence on prediction in audiovisual temporal processing during apparent motion perception. Therefore, temporal processing during apparent motion perception may result from stimulus-driven effects rather than expectancy-driven effects. On the other hand, in Section 2, although a sensory stimulus was followed by another of the same type of sensory stimulus, audiovisual temporal processing during apparent motion perception differed from that during nonapparent motion perception. This suggests the possibility that despite the stimulus-driven effects with the same type of sensory stimuli, the difference between motion and non-motion perceptions influenced temporal order perception and the window of temporal integration in audiovisual processing.

Specifically, the pathway of motion and non-motion processing may differ in audiovisual temporal perception, and there may exist some activation for determining the pathway of motion and non-motion processing by bottom-up signals. Many researchers have claimed that audio-visual stimuli are integrated at an early processing stage (Fendrich and Corballis, 2001; Bruns and Getzmann, 2008). Fendrich and Corballis (2001) suggested that the sensory capture phenomenon may be connected with low-level intersensory linking processes and it seems that auditory driving or auditory dominance depends on such low level sensory linking processes. Also, Bruns and Getzmann (2008) also reported that their findings are consistent with a low-level audiovisual integration between visual apparent motion and a single sound. Moreover, in recent years, with the growing interest in multisensory properties of motion, some researchers have raised a possibility that the area MT, which is earlier cortical area, playing an important role in visual motion processing is engaged in the audiovisual processing (Olson, et al., 2002; Kafaligonul and Stoner, 2010; Calvert, et al., 1999; Born and Bradley, 2005). In this study, the same property on temporal perception appears between predictable motion (Chapter 2) and unpredictable motion (Section 2 in Chapter 3) on audiovisual processing. Therefore, we can conclude that the pathway for motion and non-motion processing is determined by the bottom-up signals regardless of prediction as top-down factors. This fact may lead us to the mechanism which decides the pathway of motion or non-motion processing on audiovisual temporal perception by bottom-up signals.

5.5 Motion and Binding Property on Simultaneity Perception

In section 1, it became clear that the result obtained under the apparent motion condition was unaffected by the amount of flash stimulation. Besides, in section 2, apparent motion was equivalently processed regardless of prediction and the prediction as top-down factors have no effect on the results of Chapter 2. Therefore, the amount of flash stimulation and prediction have no effect on the results of the apparent motion condition. With respect to the mechanisms, Section 3 in Chapter 3 was focused on the peculiar motion perception mechanisms in humans (Watson, et al., 1986; Watson, et al., 1985). When two discrete stimuli are presented at appropriate spatiotemporal intervals, we can perceive the stimuli as a continuous motion. In other words, two discrete stimuli separated by greater intervals than 300 ms are perceived as successive stimuli, whereas if the intervals are within a range of 50 to 150 ms, the stimuli are perceived as one moving object (Getzmann, 2007; Harrar, et al., 2008; Strybel, et al., 1990). This phenomenon suggests that visual motion perception mechanisms have a binding property that stimuli are perceived as a single bounded object with spatiotemporal continuity, and the single moving object is perceived only at certain intervals (Dawson, 1991; Treisman, 1996; Kahneman, et al., 1992). Apparent motion is a most suitable phenomenon to show motion binding property in human perceptual system. Therefore, the influence of motion binding property is investigated in Section 3 in Chapter 3.

As a result, the PSS shifted toward a sound-lead stimulus and especially became closer to zero in the motion binding condition, and the JND in the motion binding condition was smaller than that in the non-motion binding condition. This shows that motion binding shifted simultaneity toward physical simultaneity (i.e., zero), and reduces the temporal window in audiovisual integration. In previous studies, the characteristics of perceptual binding in cross-modality have been reported. For example, unity assumption well represents the multi-modal binding (Bedford, 2001; Radeau and Bertelson, 1977; Bertelson and Aschersleben, 1998; Vatakis and Spence, 2007; Welch, 1999; Welch and Warren, 1980). It shows that different sensory information is bound and perceived as a same event. In particular, it has been reported that one sensory information is captured by another sensory information (Gebhard and Mowbray, 1959; Fendrich and Corballis, 2001; Bertelson and Aschersleben, 2003; Morein-Zamir, et al., 2003; Freeman and Driver, 2008; Recanzone, 2003; Vroomen and de Gelder, 2004). For example, temporal ventriloquism effect, when there is a temporal separation in audiovisual stimulation, temporal events of the visual stimuli are bound by the temporal events of the auditory stimuli (Bertelson and Aschersleben, 2003; Morein-Zamir, et al., 2003). However, there is no study that have investigated the effect of unisensory binding on the multisensory binding. In this study, we investigated the effect of motion binding property on audiovisual simultaneity. As a results, motion binding property affected audiovisual simultaneity and contributes to accurate perceptions of temporal events.

Although the difference in the PSS was shown in motion binding and non-motion binding condition, there was no difference in JNDs between motion binding condition and the SOA of 40 ms in non-motion binding condition. With respect to visual attention, it has been reported that there are two types of visual attention such as exogenous attention and endogenous attention. Exogenous attention is an involuntary system, which rises and decays quickly and acts on maximally about 100-120 ms and is effective up to 300 ms (Muller and Rabbitt, 1989; Hein, et al., 2006; Ling and Carrasco, 2006; Liu, et al., 2007; Nakayama and Mackeben, 1989; Remington, et al., 1992). On the other hand, endogenous attention is a voluntary system that deployed at more than 300 ms (Muller and Rabbitt, 1989; Hein, et al., 2006; Ling and Carrasco, 2006; Liu, et al., 2007; Nakayama and Mackeben, 1989; Remington, et al., 1992). In the results of Section 3, there is a possibility that exogenous attention by a short temporal interval is activated in motion binding perception and endogenous attention by long temporal intervals is allocated in the successive perception. Therefore, there was a need to compare between motion binding perception and non-motion binding perception by a short temporal interval. As a result, exogenous attention by a short temporal interval resulted in small JND, but unaffected the PSS. Therefore, it revealed that motion binding perception contribute to the accurate perceptions of temporal orders in audiovisual processing.

5.6 Motion and Simultaneity Perception in Human Communication

Simultaneity perception has been shown in a variety of perceptual scales as well as simple stimuli. The characteristics of simultaneity perception is differed by the perceptual scales (Zampini, et al., 2003; Wassenhove, et al., 2007). In particular, human communication shows the most high-order and complex perceptual scale. In human communication, verbal and nonverbal information is transmitted through the auditory and visual channels including the various meaning for communication (Richmond et al, 2008). Interestingly, body movement synchronization as simultaneity between two or multiple person has been shown in human communication. In this regard, the most important point is that body movement synchronization is based on simultaneity perception, which is a process perceiving body movements of each other as visual motion. In other words, we perceive the visual motion information each other, and behave on the basis of the visual motion perception.

In Chapter 4, we examined the influence of visual motion on simultaneity perception through body movement synchronization. To confirm the influence of visual motion, we detected body movement synchronization in direct face-to-face and remote communication settings using a newly-devised method of detecting nonverbal synchronization. As a result, although the mean phase difference in synchronized head nods did not differ significantly between direct face-to-face and remote communications, there were significant differences in the density, the SD and the kurtosis in the phase difference distribution of head nod synchronization between direct face-to-face and remote communications.

Our findings clarified some differences in body movement synchronization between direct face-to-face and remote communications. From a theoretical viewpoint, synchronization phenomena are based on the interaction of oscillators (Kuramoto, 1984; Acebron, et al., 2005). In the present study, the teachers presented the same auditory information to the students between the face-to-face and remote communication conditions. However, the conditions differed in the visual modality, because the teachers could see the students' back-channel signals in the direct face-to-face communication experiment but this information was not available to the teachers in the remote communication condition. Thus, the absence of visual feedback in the remote communication condition led to a lower density of synchronized head nods and a large spread with a small kurtosis in the phase difference distribution compared with those of the face-to-face communication condition. Therefore, these differences clearly showed the influence of visual motion as the mechanism of body movement synchronization in human communication. This motion feedback loop in the visual channel will play an important role in the elucidation of the synchronization mechanism in face-to-face communication and the application of remote communication.

With respect to a newly-devised method of detecting nonverbal synchronization, the findings in Chapter 4 showed a narrower temporal window for nonverbal synchronization compared with a previous study (Komori and Nagaoka, 2010). Simple reaction time, which participants press a button when light or sound is presented, was approximately 160 ms for an auditory stimulus and 190 ms for a visual stimulus (Kosinski, 2013). However, when the stimulation was complicated such as a letter or symbol, the reaction time increased within a range of 380 to 420 ms (Laming, 1968; Luce, 1986; Sternberg, 1969; Kosinski, 2013). With regard to head nods, Watanabe (1985) reported that the phase difference in head nod synchronization was 700 ms to 1400 ms, when the voice of the speaker affects the nodding of the listener. In addition, the phase difference in synchronization was approximately 800 ms to 2.0 seconds when the nodding of the listener affects the voice of the speaker. With regard to nonverbal synchronization, Komori (2010) reported that the time lag between the amplitude of the body movements of a counselor and a client occurred within a range of ± 1000 ms. In particular, there was a tendency for the counselors' body movements to be delayed by 500 ms compared with those of clients. However, in this study, the mean phase difference was 110 ms, and the SD was 320 ms. We obtained this result with a newly-devised method of detecting nonverbal synchronization based on a measure of phase difference with a high degree of temporal resolution, which was defined as the phase difference in the peak amplitudes between rhythmic body movements. Thus, this method can detect the accurate temporal properties of nonverbal synchronization.

Although we used phase difference to detect body movement synchronization, we need further indicators to characterize nonverbal synchronization properly. In the field of cognitive psychology, the point of subjective simultaneity (PSS) is commonly used, which is an indicator of subjective simultaneity in multisensory processing by a human perceptual system. The PSS is obtained by the mean of a distribution of simultaneous responses, and it has been reported that the PSS differs from physical simultaneity in multisensory integration [33–40]. Interestingly, the present study indicates that the mean phase difference in nonverbal synchronization has the same tendency as the PSS, as the mean phase difference (corresponding to the stimulus onset asynchrony at the PSS) was not zero (i.e., physically perfect synchronization). Thus, the mean phase difference can be called “the point of subjective synchronization”, and it can be used as an indicator of subjective synchronization in human communication. Another indicator in multisensory integration (in addition to the PSS) is the temporal window of integration [33–35, 40–43]. The temporal window of integration is calculated as the standard deviation (SD) of a distribution of simultaneous responses. Therefore, the SD of the phase difference distribution in nonverbal synchronization can be regarded as “the temporal window of synchronization” for evaluating subjective synchronization in human communication.

The point of subjective synchronization in nonverbal communication was shifted to a greater extent than that in multisensory integration (see Table 4-1). In addition, as mentioned above, the temporal window of integration in audiovisual speech was larger than that in an experiment using a simple pair of sound and flash stimuli (Zampini, et al., 2005; Wassenhove, et al., 2007), (see Table 4-1). This means that conversation attributes include the integration of audiovisual information. In this study, we investigated nonverbal synchronization in audiovisual speech with body movements. We found that the point of subjective synchronization was approximately 107 ms and the temporal window of synchronization was approximately 301 ms. Consequently, the point of subjective synchronization in nonverbal communication and the temporal window of synchronization are larger than those of multisensory integration in audiovisual speech. This may be attributable to the effect of body movements as an additional factor in multisensory integration in audiovisual speech.

However, from an experimental viewpoint, two factors lead to such differences between direct face-to-face and remote communications. The first factor is the spatial separation in remote communication (Morris, 2004; Jones, 1993). Although face-to-face communication between partners occurs in the same room, remote communication is carried out in different rooms. This shared space in the former condition may contribute to the differences in body movement synchronization between direct face-to-face and remote communications. The second factor is the lack of authenticity in remote communications (van der Kleij, et al., 2005; van der Kleij, et al., 2009). Remote communications have been developed to approximate face-to-face communication, but the reality created by such interactions still falls short of face-to-face communication. Therefore, the lack of authenticity in remote communications may have an effect on body movement synchronization. These factors should be studied to elucidate the mechanisms of face-to-face and remote communications and we believe that these findings will prompt further research on nonverbal synchronization in face-to-face and remote communications, including different cultures and languages.

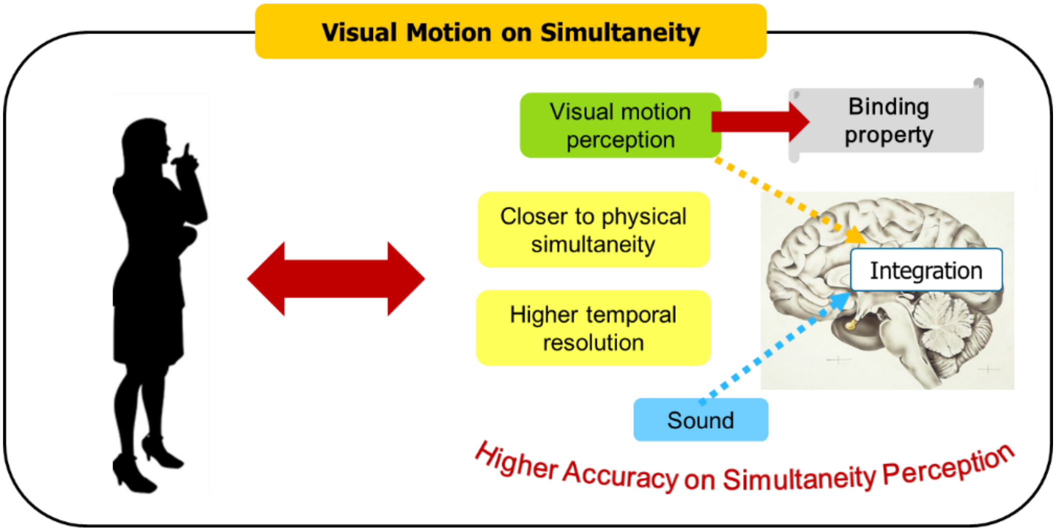


Figure 5.1. The effect of visual motion on simultaneity perception.

Chapter 6

Conclusion

Motion perception is a fundamental tool for interaction with a dynamic environment. When considering real-time interaction, temporal information of motion information is a significant factor for smooth interaction with the dynamic environment. In particular, simultaneity perception is an important key for flexible human behavior with the dynamic environment. In other words, real time interaction with the dynamic environment is generated by the relationship between motion perception and simultaneity perception. This dissertation addresses the influence of visual motion information on simultaneity perception. This dissertation's findings showed that visual motion information contributes to higher accuracy on simultaneity perceptions of temporal events in the dynamic environment. Therefore, motion perception allows flexible interaction with the dynamic environment in real time.

Bibliography

Acebron J.A., Bonilla L.L., Perez Vicente C.J., Ritort F. and Spigler, R. (2005) The Kuramoto model: a simple paradigm for synchronization phenomena. Rev Mod Phys 77, pp.137–185.

Adelson E.H. and Bergen J.R. (1985) Spatiotemporal energy models for the perception of motion, J. Opt. Soc. Amer. A 2, pp.284-299.

Alink A., Singer W. and Muckli L. (2008) 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.

Allik J. and Kreegipuu K. (1998) Multiple visual latency. Psychological Science 9, pp. 135–138.

Altmann U., Hermkes R. and Alisch L.M. (2007) Analysis of nonverbal involvement in dyadic interactions. In: Esposito A, Faundez-Zanuy M., Keller E., and Marinaro M., editors. Verbal and nonverbal communication behaviours. Berlin, Heidelberg: Springer. pp. 37–50.

Angles J., Nagatomi A. and Nakayama M. (2000) Japanese responses hai, ee and un: yes, no, and beyond. Language and Communication 20, pp.55–86.

Anstis, S.M. and Mackay, D.M. (1980) The Perception of Apparent Movement. Phil. Trans. R. Soc. Lond. B 290, pp.153–168.

Arbib M. (2010) Mirror system activity for action and language is embedded in the integration of dorsal and ventral pathways. Brain Lang 112, pp.12–24.

Bair W. and Koch C. (1996) Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. Neural Computation, 8, pp.1185–1202.

Bargh J.A. (1989) Conditional automaticity: varieties of automatic influence in social perception and cognition. In: Uleman J.S. ,and Bargh J.A., editors. Unintended thought. New York: Guilford Publications. pp. 51–69.

Bargh J.A. and Chartrand T.L. (1999) The unbearable automaticity of being. Am Psychol 54, pp.462–479.

Bavelas J.B., Coates L. and Johnson T. (2002) Listener responses as a collaborative process: The role of gaze. J Commun 52, pp.566–580.

Beauchamp M.S. (2005) See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. Curr Opin Neurobiol, 15, pp.1–9.

Bedford F.L. (2001) Towards a general law of numerical/object identity. Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 20, pp.113-175.

Bermant R.I. and Welch R.B. (1976) Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition. Percept Mot Skills 42(43), pp.487–493.

Bernieri F. (1988a) Coordinated movement and rapport in teacher–student interactions. Journal of Nonverbal Behavior, 12(2), pp.120–138.

Bernieri F., Reznick J.S., Rosenthal R. (1988b) Synchrony, pseudo-synchrony, and dissynchrony: Measuring the entrainment process in mother–infant interactions. Journal of Personality and Social Psychology, 54(2), pp.243–253.

Bernieri F.J., Gillis J.S., Davis J.M. and Grahe J.G. (1996) Dyad rapport and accuracy of its judgment across situations: a lens model analysis. J Pers Soc Psychol 71, pp.110–129.

Bertelson P. and Aschersleben G. (1998) Automatic visual bias of perceived auditory location. Psychonomic Bulletin & Review, 5, pp.482-489.

Bertelson P. and Aschersleben G. (2003) Temporal ventriloquism: crossmodal interaction on the time dimension. 1. Evidence from auditory–visual temporal order judgment. International Journal of Psychophysiology 50, pp.147–155.

Bertelson P. and Radeau M. (1981) Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. Percept Psychophys29(6):578–584.

Born R.T. and Bradley D.C. (2005) Structure and function of visual area MT. Annual Review of Neuroscience, 28, pp.157–189.

Braddick, O.J., Ruddock, K. H., Morgan, M.J. and Marr, D. (1980) Low-Level and High-Level Processes in Apparent Motion. Philosophical transactions Royal Society London B 290, pp.137-151.

Brenner E. and Smeets J.B.J. (2000) Motion extrapolation is not responsible for the flash-lag effect. Vis Res;40, pp.1645–1648.

Briggs R. and Perrott D.R. (1972) Auditory apparent movement under dichotic listening conditions. Journal of Experimental Psychology 92, pp.83–91.

Bruns P. and Getzmann S. (2008) Audiovisual influences on the perception of visual apparent motion: exploring the effect of a single sound. Acta Psychol 129(2), pp.273–283.

Buccino G., Binkofski F. and Riggio L. (2004) The mirror neuron system and action recognition. Brain Lang., 89 (2004), pp.370–376.

Bueti D., Bahrami B. and Walsh, V. (2008) Sensory and association cortex in time perception. Journal of Cognitive Neuroscience, 20, pp.1054–1062.

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

Burr, D. and Thompson, P. (2011) Motion psychophysics: 1985–2011. Vision Research, 51(13), pp.1431–1456.

Burt, P. and Sperling, G. (1981) Time, distance, and feature trade-offs in visual apparent motion. Psych. Rev. 88, pp.171–195.

Busse L., Katzner S. and 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, pp.16380–16385.

Callaway E.M. (2005) Structure and function of parallel pathways in the primate early visual system. Journal of Physiology;566(1), pp.13–19.

Calvert G.A., Brammer M.J., Bullmore E.T., Campbell R., Iversen S.D. and David, A.S. (1999) Response amplification in sensory-specific cortices during crossmodal binding. Neuroreport, 10, pp.2619–2623.

Casagrande V. and Norton T. (1991) Lateral geniculate nucleus: a review of its physiology and function. in The Neural Basis of Visual Function, ed Leventhal A. G. (MacMillan, London), pp.41–73.

Cesar P., Chorianopolous K. and Jensen J.J. (2008) Social television and user interaction. ACM Comput. Entertain. 6(1), pp.1–10.

Ciaramitaro V.M., Buračas G.T. and Boynton G.M. (2007) Spatial and Cross-Modal Attention Alter Responses to Unattended Sensory Information in Early Visual and Auditory Human Cortex. Journal of Neurophysiology 98(4), pp.2399-2413.

Cold reading: Wikipedia article (in Japanese). Available: http://en.wikipedia.org/wiki/Cold_reading. Accessed: 17 September 2012.

Condon W.S. and Sander L.W. (1974) Synchrony demonstrated between movements of the neonate and adult speech. Child Dev 45, pp.456–462.

Coull J.T. and Nobre A.C. (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, pp.7426–7435.

Darwin C., Ekman P. and Prodger P. (1998) The expression of the emotions in man and animals, 3rd edition. London: Harper Collins. 473 p.

Dawson M. (1991) The how and why of what went where in apparent motion: modeling solutions to the motion correspondence problem. Psychological Review 98, pp.569–603.

Duncan J. (1980) The demonstration of capacity limitation. Cognitive Psychology 12, pp.75-96.

Eagleman D.M. (2008) Human time perception and its illusions. Curr Opin Neurobiol 18, pp.131–136.

Easton R.D. and Basala M. (1982) Perceptual dominance during lipreading. Perception & Psychophysics, 32, pp.562-570.

Fang F. and He S. (2005) Cortical responses to invisible objects in the human dorsal and ventral pathways Nat. Neurosci. 8, pp.1380–1385.

Felleman D.J. and Van Essen D.C. (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex 1, pp.1–47.

Fendrich R. and Corballis P.M. (2001) The temporal cross-capture of audition and vision. Perception & Psychophysics 63, pp.719–725.

Finney D.J. (1952) Probit analysis: A statistical treatment of the sigmoid response curve. London: Cambridge University Press. 318 p.

Freeman E. and Driver J. (2008) Direction of visual apparent motion driven solely by timing of a static sound. Current Biology 18, pp.1262–1266.

Frissen I., Ziat M., Campion G., Hayward V. and Guastavino C. (2012) The effects of voluntary movements on auditory-haptic and haptic-haptic temporal order judgments. Acta Psychol (Amst) 141, pp.140–148.

Gebhard J.W. and Mowbray G.H. (1959) On discriminating the rate of visual flicker and auditory flutter. Am J Psychol 72, pp.521–529.

Getzmann S. (2007) The effect of brief auditory stimuli on visual apparent motion. Perception 36, pp.1089–1103.

Ghazanfar A.A. and Schroeder C.E. (2006) Is neocortex essentially multisensory? Review TRENDS in Cognitive Sciences Vol.10 No.6, pp.278-285.

Goodale M.A. and Westwood D.A. (2004) An evolving view of duplex vision: separate but interacting cortical pathways for perception and action Curr Opin Neurobiol, 14, pp. 203–211.

Green K.P. and Gerdeman A. (1995) Cross-modal discrepancies in coarticulation and the integration of speech information: The McGurk effect with mismatched vowels. Journal of Experimental Psychology: Human Perception & Performance, 21, pp.1409-1426.

Green K.P., Kuhl P.K., Meltzoff A.N. and Stevens E.B. (1991) Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. Perception & Psychophysics, 50, pp.524-536.

Grondin S. (2010) Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. Atten Percept Psychophys 72, pp.561–582.

Grossberg, S. and Rudd, M.E. (1992) Cortical dynamics of visual motion perception: short-range and long-range apparent motion. Psychological Review 99, pp.78–121.

Guski R. and Troje N.F. (2003) Audiovisual phenomenal causality. Perception & Psychophysics, 65, pp.789-800.

Hadar U., Steiner T.J., Grant E.C. and Clifford R.F. (1984) The timing of shifts of head postures during conversation. Hum Mov Sci 3, pp.237–245.

Hairston W.D., Wallace M.T., Vaughan J.W., Stein B.E., Norris J.L. and Schirillo J.A. (2003) Visual localization ability influences cross-modal bias. Journal of Cognitive Neuroscience, 15, pp.20-29.

Harboe G., Massey N., Metcalf C., Wheatley D. and Romano G. (2008) The uses of social television. ACM Comput. Entertain. 6(1):(8), pp.1–15.

Harrar V., Winter R. and Harris L.R. (2008) Visuotactile apparent motion. Perception & Psychophysics 70, pp.807–817.

Hein E., Rolke B. and Ulrich R. (2006) Visual attention and temporal discrimination: Differential effects of automatic and voluntary cueing. Visual Cognition, 13(1), pp.29-50.

Hendry S.H.C. and Reid R.C. (2000) The koniocellular pathway in primate vision. Annu. Rev. Neurosci. 23, pp.127–153.

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

Hilgetag C.-C., O'Neill M.A. and Young M.P. (1996) Indeterminate organization of the visual system. Science 271, pp.776–777.

Hodgins H. and Zuckerman M. (1990) The effect of nonverbal sensitivity on social interaction. J Nonverbal Behav 14, pp.155–170.

Iacoboni M. and Dapretto M. (2006) The mirror neuron system and the consequences of its dysfunction Nat. Rev. Neurosci.7, pp.942–951.

Ivry R.B. and Schlerf J.E. (2008) Dedicated and intrinsic models of time perception. Trends Cogn Sci 12, pp.273–280.

Jack C.E. and Thurlow W.R. (1973) Effects of degree of visual association and angle of displacement on the “ventriloquism” effect. Perceptual & Motor Skills, 37, pp.967–979.

Jakowśki P. and Verleger R. (2000) Attentional bias toward low intensity stimuli: an explanation for the intensity dissociation between reaction time and temporal order judgment? Consciousness & Cognition 9, pp.435–456.

Jakowśki P., Jaroszyk F. and Hojan-Jezierska D. (1990) Temporal-order judgments and reaction time for stimuli of different modalities. Psychol Res 52, pp.35–38.

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

Jones S. (1993) A sense of space: virtual reality, authenticity and the aural. Critical Studies in Mass Communication 10, pp.238–252.

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

Kahneman D., Treisman A. and Gibbs B. (1992) The reviewing of object files: object-specific integration of information. Cogn Psycho 24, pp.175–219.

Kanabus M., Szelg E., Rojek E. and Pöppel E. (2002) Temporal order judgment for auditory and visual stimuli. Acta Neurobiol. Exp 62, pp.263–270.

Kayser C., Petkov C.I. and Logothetis N.K. (2008) Visual modulation of neurons in auditory cortex. Cerebral Cortex 18, pp.1560–1574.

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

King A.J. (2005) Multisensory integration: Strategies for synchronization. Current Biology, 15:R339-R341.

Kita S. and Ide S. (2007) Nodding, aizuchi, and final particles in Japanese conversation: how conversation reflects the ideology of communication and social relationships. J Pragmat 39, pp.1242–1254.

Klein R.M. (1977) Attention and visual dominance: A chronometric analysis. Journal of Experimental Psychology: Human Perception & Performance 3, pp.365–378.

Komori M. and Nagaoka C. (2010) The relationship between body movements of clients and counselors in psychotherapeutic counseling: a study using the video-based quantification method. Japanese Journal of Cognitive Psychology 8, pp.1–9.

Kosinski R.J. (Sept 2013) A literature review on reaction time.

Krekelberg B. (2008) Motion detection mechanisms. In T. D. Albright & R. Masland (Eds.), The senses: A comprehensive reference (vol. 2, pp. 133–154). Oxford, UK: Elsevier.

Kuramoto Y. (1984) Chemical oscillations, waves and turbulence. Springer-Verlag, Berlin. 158 p.

Kwon J., Ogawa K. and Miyake Y. (2014) The effect of visual apparent motion on audiovisual simultaneity. PLoS One 9(10): e110224. pp.1–9

Lakin J.L. (2006) Automatic cognitive processes and nonverbal communication. In: Manusov V. and Patterson M.L., editors. The Sage handbook of nonverbal communication. Thousand Oaks: Sage Publications. pp.59–77.

Laming D.R.J. (1968) Information theory of choice-reaction times. London: Academic Press.

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

Larsen A., Farrell J.E. and Bundesen C. (1983) Short- and Long-Range Processes in Visual Apparent Movement," Psychol Res 45, pp.11–18.

Laurienti P.J., Kraft R.A., Maldjian J.A., Burdette J.H. and Wallace M.T. (2004) Semantic congruence is a critical factor in multisensory behavioral performance. Experimental Brain Research, 158, pp.405-414.

Lewald J. and Guski R. (2003) Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. Cognitive Brain Research 16, pp.468–478.

Ling S. and Carrasco M. (2006) Sustained and transient covert attention enhance the signal via different contrast response functions. Vision Research, 46(8-9), pp.1210-1220.

Liu T., Stevens S.T. and Carrasco M. (2007) Comparing the time course and efficacy of spatial and feature-based attention. Vision Research, 47(1), pp.108-113.

Luce R.D. (1986) Response times: Their role in inferring elementary mental organization. New York: Oxford University Press.

MacDorman, C.E. (1962) Synchronization with auditory models of varying complexity. Perceptual and Motor Skills, 15, pp.595–602.

Manning M.L., Finlay D.C. and Fenelon B. (1988) Visual evoked potentials to stimuli in apparent motion. Vision Research, 28, pp.965–974.

Martin K.A.C. (1992) Parallel pathways converge. Curr. Biol. 2, pp.555–557.

Mattes S. and Ulrich R. (1998) Directed attention prolongs the perceived duration of a brief stimulus. Perception & Psychophysics 60, pp.1305–1317.

Mauk M.D. and Buonomano D.V. (2004) The neural basis of temporal processing. Annual Review of Neuroscience 27, pp.307–340.

Maynard S.K. (1987) Interactional function of a nonverbal sign—Head movement in Japanese dyadic casual conversation. J Pragmat 11, pp.589–606.

Maynard S.K. (1990) Conversation management in contrast: listener response in Japanese and American English. Journal of Pragmatics 14, pp.397–412.

Merigan W.H. and Maunsell J.H.R. (1993) How parallel are the primate pathways? Annu. Rev. Neurosci. 16, pp.369–402.

Miyake Y., Onishi Y. and Pöppel E. (2004) Two types of anticipation in synchronization tapping. Acta Neurobiologiae Experimentalis, vol.64, pp.415–426.

Miyake Y., Onishi Y. and Pöppel E. (2007) Two types of anticipatory-timing mechanism in synchronization tapping. In N Osaka, I Rentschler, I Biederman (Eds.), Object Recognition, Attention, and Action, Springer-Verlag, Tokyo, pp.231–244.

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

Morris D. (2004) The sense of space. Albany, NY: SUNY Press. 232 p.

Muller H.J. and Rabbitt P.M. (1989) Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. Journal of Experimental Psychology: Human Perception and Performance, 15(2), pp.315–330.

Nagaoka C. and Komori M. (2008) Body movement synchrony in psychotherapeutic counseling: a study using the video-based quantification method. IEICE Trans Inf Syst E91-D, pp.1634–1640.

Nakayama K. and Mackeben M. (1989) Sustained and transient components of focal visual attention. Vision Research, 29(11), pp.1631-1647.

Nassi J.J. and Callaway E.M. (2009) Parallel processing strategies of the primate visual system Nat. Rev. Neurosci., 10, pp.360–372.

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

Nijhawan R. (1994) Motion extrapolation in catching. Nature 370, pp.256–257.

Nijhawan R. (2002) Neural delays, visual motion and the flash-lag effect. TRENDS in Cognitive Sciences Vol.6 No.9, pp.387–393.

Nishi A., Yokoyama M., Ogawa K., Ogata T., Nozawa T. and Miyake Y. (2014) Effects of voluntary movements in audio-tactile temporal order judgment. IEICE Transactions on Information and Systems, Vol.E97-D, No.6, pp.1567-1573.

Nishida S. (2011) Advancement of motion psychophysics: Review 2001–2010. Journal of Vision, 11 (5): 11, pp. 1–53.

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

Online Stopwatch website. Available: http://www.online-stopwatch.com/. Accessed: 15 October 2012.

Oztop E., Kawato M., Arbib M.A. (2012) Mirror neurons: functions, mechanisms and models. Neurosci. Lett. 540, pp. 43–55.

O’Craven K.M., Rosen B.R., Kwong K.K., Treisman A. and Savoy R.L. (1997) Voluntary attention modulates fMRI activity in human MT-MST. Neuron 18: pp. 591–598.

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

Paxton A., Dale R. (2013) Frame-differencing methods for measuring bodily synchrony in conversation. Behavior Research Methods, 45(2), pp.329–343.

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

Posner M.I., Nissen M.J. and Klein R.M. (1976) Visual dominance: an information-processing account of its origins and significance. Psychological Review 83, pp.157–171.

Posner M.I., Snyder C.R. and Davidson B.J. (1980) Attention and the Detection of Signals. Journal of Experimental Psychology 109, pp.160–174.

Pöppel E. (1988) Mindworks: Time and conscious experience. New York: Harcourt Brace Jovanovich. 221 p.

Radeau M. (1994) Auditory-visual spatial interaction and modularity. Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 13, pp.3–51.

Radeau M. and Bertelson P. (1977) Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations. Perception & Psychophysics, 22, pp.137–146.

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

Ramseyer F. and Tschacher W. (2011) Nonverbal synchrony in psychotherapy: coordinated body-movement reflects relationship quality and outcome. J Consult Clin Psychol 79, pp.284–295.

Recanzone G.H. (2003) Auditory influences on visual temporal rate perception, Journal of Neurophysiology 89, pp.1078–1093.

Remington R.W., Johnston J.C. and Yantis S. (1992) Involuntary attentional capture by abrupt onsets. Perception & Psychophysics, 51(3) , pp.279-290.

Repp B.H. (2005) Sensorimotor synchronization: A review of the tapping literature. Psychonomic Bulletin & Review, 12, pp.969–992.

Repp B.H. and Keller P. (2008) Sensorimotor synchronization with adaptively timed sequences. Human Movement Science. 27, pp.423–456.

Richmond V.P., McCroskey J.C. and Hickson M.L. (2008) Nonverbal behavior in interpersonal relations. Allyn & Bacon, Boston: Pearson Education. 366 p.

Rizzolatti G. and Craighero L. (2004) The mirror-neuron system. Annu Rev Neurosci. 27, pp.169–192.

Rizzolatti G. (2005) The mirror neuron system and its function in humans. Anat Embryol 210, pp.419–421.

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

Sanabria D., Soto-Faraco S., Chan J.S. and Spence C. (2004) When does visual perceptual grouping affect multisensory integration? Cognitive, Affective, & Behavioral Neuroscience, 4, pp.218-229.

Scheef L., Boecker H., Daamen M., Fehse U., Landsberg M.W., Granath, D.O., et al. (2009) Multimodal motion processing in area V5/MT: Evidence from an artificial class of audio visual events. Brain Research, 1252, pp.94–104.

Scheier C.R., Nijhawan R. and Shimojo S. (1999) Sound alters visual temporal resolution. Investigative Ophthalmology & Visual Science, 40, p.4169.

Schlag J. and Schlag-Rey M. (2002) Through the eye slowly: delays and localization errors in the visual system. Nat. Rev. Neurosci. 3, pp.191–200.

Schmidt RC, Carello C and Turvey MT (1990) Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. J Exp Psychol Hum Percept Perform 16(2), pp.227–247.

Schneider K.A. and Bavelier D. (2003) Components of visual prior entry. Cognitive Psychology 47, pp.333–366.

Schneider W. and Chein J.M. (2003) Controlled & automatic processing: behavior, theory, and biological mechanisms. Cogn Sci 27, pp.525–559.

Shadlen M.N., Zohary E., Britten K.H. and Newsome W.T. (1993) Directional properties of MT neurons examined with motion energy filtered apparent motion stimuli. Invest. Ophthalmol. Vis. Sci. Suppl. 34(4), pp.908.

Shimono M., Mano H. and Niki K. (2012) The brain structural hub of interhemispheric information integration for visual motion perception Cereb. Cortex, 22, pp.337–344.

Simoncelli E.P. and Heeger D.J. (1998) A model of neuronal responses in visual area MT. Vision Res. 38(5), pp.743–61.

Slutsky D.A. and Recanzone G.H. (2001) Temporal and spatial dependency of the ventriloquism effect. NeuroReport, 12, pp.7–10.

Spence C. and Driver J. (1997) On measuring selective attention to an expected sensory modality. Perception & Psychophysics 59, pp.389–403.

Spence C. and Parise C. (2010) Prior-entry: a review. Consciousness and Cognition 19, pp.364–379.

Spence C. and Squire S. (2003) Multisensory integration: Maintaining the perception of synchrony. Current Biology, 13, pp.R519–R521.

Spence C., Nicholls M.E.R. and Driver J. (2001) The cost of expecting events in the wrong sensory modality. Perception & Psychophysics 63, pp.330–336.

Spence C., Sanabria D. and Soto-Faraco S. (2007) Intersensory Gestalten: Assessing the influence of intramodal perceptual grouping on crossmodal interactions. In K. Noguchi (Ed.), The psychology of beauty and Kansei: New horizons of Gestalt perception, pp. 519–579.

Spence C., Shore D.I. and Klein R.M. (2001) Multisensory prior entry. Journal of Experimental Psychology 130, pp.799–832.

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

Sternberg S. (1969) Memory scanning: Mental processes revealed by reaction time experiments. Am Sci 57: 421–457.

Stivers T. (2008) Stance, alignment and affiliation during story telling: When nodding is a token of affiliation. Research on Language and Social Interaction 41(1), pp.31-57.

Stone J., Dreher B. and Leventhal A. (1979) Parallel and hierarchical mechanisms in the organization of visual cortex. Brain Res. Rev. 1, pp.345–394.

Strybel T.Z., Manligas C.L., Chan O. and Perrott D.R. (1990) A comparison of the effects of spatial separation on apparent motion in the auditory and visual modalities. Perception & Psychophysics 47, pp.439–448.

Tickle-Degnen L. and Rosenthal R. (1990) The nature of rapport and its nonverbal correlates. Psychol Inq. 1, pp.285–293.

Treisman A. (1996) The binding problem. Curr Opin Neurobiol 6, pp.171–178.

Treue S. and Martinez Trujillo J.C. (1999) Feature-based attention influences motion processing gain in macaque visual cortex. Nature 399, pp.575–579.

Vaina L. M. and Soloviev S. (2004) First-order and second-order motion: Neurological evidence for neuroanatomically distinct systems. Progress in Brain Research, 144, pp.197–212.

van der Kleij R., Paashuis R.M. and Schraagen J.M.C. (2005) On the passage of time: temporal differences in video-mediated and face-to-face interaction. International Journal of Human-Computer Studies 62, pp.521–542.

van der Kleij R., Schraagen J., Werkhoven P. and De Dreu C. (2009) How conversations change over time in face-to-face and video-mediated communication. Small Group Research 40(4), pp.355−381.

Van Essen D.C., Anderson C.H. and Felleman, D.J. (1992) Information processing in the primate visual system: an integrated systems perspective. Science 255, pp.419–423.

Vatakis A. and Spence C. (2006) Audiovisual synchrony perception for speech and music using a temporal order judgment task. Neuroscience Letters, 393, pp.40-44.

Vatakis A. and Spence C. (2007) Crossmodal binding: evaluating the “unity assumption” using audiovisual speech stimuli Percept. Psychophys, 69, pp.744–756.

Vroomen J. and Keetels M. (2006) The spatial constraint in intersensory pairing: No role in temporal ventriloquism. Journal of Experimental Psychology: Human Perception & Performance, 32, pp.1063-1071.

Vroomen J. and de Gelder B. (2004) Temporal ventriloquism: sound modulates the flash-lag effect. Journal of Experimental Psychology: Human Perception & Performance 30, pp.513–518.

Vroomen J., Keetels M. (2010) Perception of intersensory synchrony: a tutorial review. Attention, Perception, & Psychophysics 72, pp.871–884.

Walker S., Bruce V. and O'Malley C. (1995) Facial identity and facial speech processing: Familiar faces and voices in the McGurk effect. Perception & Psychophysics, 57, pp.1124-1133.

Wassenhove V., Grant K.W. and Poeppel D. (2007) Temporal window of integration in auditory–visual speech perception. Neuro-psychologia 45, pp.598–607.

Watanabe T. (1985) An analysis of entrainment (voice–movement synchronization) in communication among adults. Journal of Information Processing (JIP) 26, pp.272–277.

Watson A.B. and Ahumada Jr.A.J. (1985) Model of human visual-motion sensing J Opt Soc Am A 2, pp.322–342.

Watson A.B., Ahumada Jr.A.J and Farrell J.E. (1983) The window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays. NASA Tech. Paper TP-2211.

Watson A.B., Ahumada Jr.A.J. and Farrell J.E. (1986) Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays. J. Opt. Soc. Am. A 3, pp.300-307.

Welch R.B. (1999) Meaning, attention, and the ''unity assumption'' in the intersensory bias of spatial and temporal perceptions. In Aschersleben G., Bachmann T. and Müsseler J. (Eds.), Cognitive contributions to the perception of spatial and temporal events. Amsterdam: Elsevier Science, pp.371–387.

Welch R.B. and Warren D.H. (1980) Immediate perceptual response to intersensory discrepancy. Psychological Bulletin, 88, pp.638–667.

White S. (1989) Backchannels across cultures: a study of Americans and Japanese. Language in Society 18, pp.59–76.

Woodrow H. (1932) The effect of rate of sequence upon the accuracy of synchronization. Journal of Experimental Psychology, 15, pp.357–379.

Wu H.R. and Rao K.R. (2005) Digital Video Image Quality and Perceptual Coding. Boca Raton, FL: CRC Press.

Yngve V.H. (1970) On getting a word in edgewise. Papers from the 6th regional meeting. Chicago Linguistic Society, Chicago, pp.567–578.

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

Zampini M., Shore D.I. and Spence C. (2003) Audiovisual temporal order judgments. Experimental Brain Research 152, pp.198–210.

Acknowledgments

First and foremost, I am deeply grateful to Prof. Yoshihiro Miyake, who is my supervisor and a great teacher. Prof. Yoshihiro Miyake has supported me throughout my dissertation with his knowledge and patience. He has shown the attitude and the substance of a genius. He continually and persuasively conveyed a spirit of adventure in regard to research and scholarship, and an excitement in regard to teaching. I was always motivated by his positive thinking, his enthusiasm, and his trust in me. He was my supervisor I respect but he was sometimes my father, colleague and friend. I also have learnt the high standard of research and the importance of social communication. Without his support and effort, this dissertation would have never been accomplished. The three years with him has been a priceless experience for me.

I am also very grateful to Assist. Prof. Ken-ichiro Ogawa, who has been a tremendous mentor for me. He has helped me in critical moments and helped me improving my research and revising my journal manuscripts. He kindly spent time checking and approving my manuscripts. He has always encouraged my research and allowed me to grow as a research scientist. Without his supervision and constant help this dissertation would not have been possible.

I would also pay tribute to the respected reviewers of my dissertation, Prof. Takao Terano, Prof. Katsumi Nitta, Prof. Kiyohiko Nakamura and Prof. Eizo Miyashita. I feel honored that you are willing to participate in my examining committee. I really want to thank you for your brilliant comments, and constructive advices and suggestions.

I would like to extend my gratitude to Assist. Prof. Taiki Ogata of the University of Tokyo. Prof. Taiki Ogata has helped me through useful discussions. He is greatly special to me because he offered helpful comments on an earlier version of this dissertation. I also would like to thank Mr. JongHwan Kim of the Korean Advanced Institute of Science and Technology for assistance in data programming.

I would like to thank everyone in Miyake Laboratory for supporting and helping me succeed my degree. In my daily work I have been blessed with a friendly and cheerful group of fellow students. All students have supported my research through

lab discussion, and meaningful and valuable life in Japan. I also would like to acknowledge Japan Tobacco Inc. for the financial support during my study in Tokyo Tech.

A special thanks to my family. Words cannot express how grateful I am my Jesus Christ, father, mother, two elder brothers and elder brother’s wives, and my beloved niece and nephew for all of the sacrifices that you’ve made on my behalf. Your prayer for me and encouragement through phone calls were what sustained me thus far. I am forever grateful. This dissertation stands as a testament to your unconditional love and encouragement. I would also like to thank all of my friends who supported me, and incented me to strive towards my goal.

Appendix

Table A-1. PSSs in Chapter 2. A paired t-test is used to compare the results between the apparent motion condition and the normal condition. The table shows the results of paired t-tests of PSSs, indicating a significant difference between the TOJ tasks in the apparent motion condition and those in the normal condition.

Participant ID	PSS		Paired t-test
	apparent motion condition	normal condition	
1	-15.32	-3.49	t(15) = –2.33, *p* = 0.034
2	-6.97	7.62	
3	19.47	29.98	
4	13.00	13.36	
5	-43.69	-10.70	
6	15.22	36.31	
7	-27.06	31.71	
8	-31.14	17.21	
9	-5.37	-13.29	
10	-19.23	60.53	
11	9.02	37.69	
12	-8.93	-13.23	
13	24.12	-22.79	
14	-41.77	-37.36	
15	27.02	37.07	
16	11.04	19.37	

Table A-2. JNDs in Chapter 2. A paired t-test is used to compare the results between the apparent motion condition and the normal condition. The table shows the results of paired t-tests of JNDs, indicating a significant difference between the TOJ tasks in the apparent motion condition and those in the normal condition.

Participant ID	**JND** **apparent motion condition**	 **normal condition**	**Paired t-test**
1	32.64	57.48	t(15) = –3.57, *p* = 0.001
2	14.60	47.22	
3	33.46	60.98	
4	17.38	34.70	
5	40.85	94.88	
6	30.05	35.65	
7	21.33	32.07	
8	55.26	55.29	
9	44.78	62.88	
10	53.03	56.23	
11	22.73	30.10	
12	12.57	31.30	
13	69.21	61.48	
14	29.69	26.27	
15	15.09	19.89	
16	23.41	31.74	

Table A-3. PSSs in Chapter 3.1. A repeated-measures analysis of variance (ANOVA) and multiple comparisons with Holm correction are used to compare the PSSs in the apparent motion condition and those in the successive condition. The table indicates a statistically significant difference between the apparent motion and successive conditions.

Participant ID	PSS			Repeated-measures analysis of variance (ANOVA) and Multiple comparisons with Holm correction
	apparent motion condition	successive condition (SOA:300ms)	successive condition (SOA:500ms)	
1	-17.06	22.76	32.31	F(2, 23) = 15.83, p < 0.001, apparent motion and an SOA of 300 ms: *p* = 0.042, apparent motion and an SOA of 500 ms: p = 0.012
2	-17.35	-1.06	3.36	
3	-3.05	-13.26	-9.29	
4	13.02	47.11	36.73	
5	-21.98	-13.83	-14.58	
6	27.44	46.56	54.40	
7	-13.17	-4.07	26.82	
8	-32.00	-20.14	-18.64	
9	21.03	19.05	31.04	
10	43.10	62.10	51.69	
11	-38.93	-31.71	-29.20	
12	-25.10	-34.59	-26.37	

Table A-4. JNDs in Chapter 3.1. A repeated-measures analysis of variance (ANOVA) and multiple comparisons with Holm correction are used to compare the JNDs in the apparent motion condition and those in the successive condition. The table indicates a statistically significant difference between the apparent motion and successive conditions.

Participant ID	JND apparent motion condition	 successive condition (SOA:300ms)	 successive condition (SOA:500ms)	Repeated-measures analysis of variance (ANOVA) and Multiple comparisons with Holm correction
1	23.03	45.81	49.79	F(2, 23) = 25.03, p < 0.001, apparent motion and an SOA of 300 ms: *p* = 0.002, apparent motion and an SOA of 500 ms: p < 0.001
2	32.22	41.55	51.87	
3	31.51	32.15	36.09	
4	20.91	36.68	38.33	
5	38.91	68.33	51.87	
6	30.09	43.53	46.81	
7	29.56	78.78	69.90	
8	34.73	59.84	66.60	
9	19.50	22.42	19.15	
10	19.95	26.64	26.46	
11	12.58	32.07	44.98	
12	23.29	38.32	40.35	

Table A-5. PSSs in Chapter 3.2. An unpaired t-test is used to compare the results of apparent motion conditions in Chapter 1 and Section 2 in Chapter 3. The table shows that unpaired t-tests of PSSs revealed no significant difference between the TOJ tasks in the predictable apparent motion condition (Chapter 1) and unpredictable apparent motion condition (Section 2 in Chapter 3).

Participant ID	PSS		unpaired t-test
	apparent motion condition (Chapter 1)	apparent motion condition (Section 2 in Chapter 2)	
1	-15.32	11.65	t(26) = –0.11, *p* = 0.92
2	-6.97	-4.12	
3	19.47	11.28	
4	13.00	-15.88	
5	-43.69	-15.02	
6	15.22	0.00	
7	-27.06	9.63	
8	-31.14	16.88	
9	-5.37	39.53	
10	-19.23	-9.40	
11	9.02	-41.71	
12	-8.93	-51.56	
13	24.12		
14	-41.77		
15	27.02		
16	11.04		

Table A-6. JNDs in Chapter 3.2. An unpaired t-test is used to compare the results of apparent motion conditions in Chapter I and Section 2 in Chapter 3. The table shows that unpaired t-tests of PSSs revealed no significant difference between the TOJ tasks in the predictable apparent motion condition (Chapter 1) and unpredictable apparent motion condition (Section 2 in Chapter 3).

Participant ID	JND apparent motion condition (Chapter 1)	 apparent motion condition (Section 2 in Chapter 2)	unpaired t-test
1	32.64	36.79	t(26) = –0.12, *p* = 0.91
2	14.60	48.56	
3	33.46	22.60	
4	17.38	41.70	
5	40.85	19.69	
6	30.05	36.79	
7	21.33	34.69	
8	55.26	16.78	
9	44.78	22.17	
10	53.03	22.92	
11	22.73	64.25	
12	12.57	28.39	
13	69.21		
14	29.69		
15	15.09		
16	23.41		

Table A-7. PSSs of Experiment 1 in Chapter 3.3. A paired t-test is used to compare the results between the motion binding condition and the simultaneous condition. The table shows the results of paired t-tests of PSSs, indicating a significant difference between the TOJ tasks in the motion binding condition and those in the simultaneous condition.

Participant ID	PSS		unpaired t-test
	motion binding condition	simultaneous condition	
1	-15.06	-11.39	t(11) = –2.43, *p* = 0.034
2	-8.93	11.01	
3	-5.02	11.96	
4	2.97	11.03	
5	17.01	23.29	
6	-13.15	39.56	
7	-23.06	-25.58	
8	27.44	39.31	
9	6.11	16.38	
10	-19.08	-7.01	
11	-35.28	49.04	
12	23.17	15.11	

Table A-8. JNDs of Experiment 1 in Chapter 3.3. A paired t-test is used to compare the results between the motion binding condition and the simultaneous condition. The table shows the results of paired t-tests of JNDs, indicating a significant difference between the TOJ tasks in the motion binding condition and those in the simultaneous condition.

Participant ID	JND		unpaired t-test
	apparent motion condition (Chapter 1)	apparent motion condition (Section 2 in Chapter 2)	
1	23.56	36.67	t(11) = –4.49, *p* = 0.0009
2	12.57	19.10	
3	24.18	46.80	
4	13.52	22.39	
5	16.06	28.80	
6	28.67	41.93	
7	13.00	32.58	
8	30.09	25.05	
9	64.50	91.12	
10	12.02	20.90	
11	52.89	80.73	
12	26.02	26.44	

Table A-9. PSSs of Experiment 2 in Chapter 3.3. A repeated-measures analysis of variance (ANOVA) and multiple comparisons with Holm correction are used to compare the PSSs in the apparent motion condition and those in the successive condition. The table indicates a statistically significant difference between the apparent motion and successive conditions.

Participant ID	PSS			Repeated-measures analysis of variance (ANOVA) and Multiple comparisons with Holm correction
	motion binding condition	successive condition (SOA:300ms)	successive condition (SOA:500ms)	
1	-17.06	22.76	32.31	F(2, 23) = 15.83, p < 0.001, motion binding and an SOA of 300 ms: *p* = 0.042, apparent motion and an SOA of 500 ms: p = 0.012
2	-17.35	-1.06	3.36	
3	-3.05	-13.26	-9.29	
4	13.02	47.11	36.73	
5	-21.98	-13.83	-14.58	
6	27.44	46.56	54.40	
7	-13.17	-4.07	26.82	
8	-32.00	-20.14	-18.64	
9	21.03	19.05	31.04	
10	43.10	62.10	51.69	
11	-38.93	-31.71	-29.20	
12	-25.10	-34.59	-26.37	

Table A-10. JNDs of Experiment 2 in Chapter 3.3. A repeated-measures analysis of variance (ANOVA) and multiple comparisons with Holm correction are used to compare the JNDs in the apparent motion condition and those in the successive condition. The table indicates a statistically significant difference between the apparent motion and successive conditions.

Participant ID	JND			Repeated-measures analysis of variance (ANOVA) and Multiple comparisons with Holm correction
	motion binding condition	successive condition (SOA:300ms)	successive condition (SOA:500ms)	
1	23.03	45.81	49.79	F(2, 23) = 25.03, p < 0.001, motion binding and an SOA of 300 ms: *p* = 0.002, apparent motion and an SOA of 500 ms: p < 0.001
2	32.22	41.55	51.87	
3	31.51	32.15	36.09	
4	20.91	36.68	38.33	
5	38.91	68.33	51.87	
6	30.09	43.53	46.81	
7	29.56	78.78	69.90	
8	34.73	59.84	66.60	
9	19.50	22.42	19.15	
10	19.95	26.64	26.46	
11	12.58	32.07	44.98	
12	23.29	38.32	40.35	

Table A-11. PSSs of Experiment 3 in Chapter 3.3. A repeated-measures analysis of variance (ANOVA) and multiple comparisons with Bonferroni correction are used to compare the PSSs in the motion binding condition and those in the non-motion binding condition. The table indicates a statistically significant difference between the motion binding and non-motion binding conditions.

Participant ID	PSS motion binding condition	 non-motion binding condition (SOA:40ms)	 non-motion binding condition (SOA:300ms)	Repeated-measures analysis of variance (ANOVA) and Multiple comparisons with Bonferroni correction
1	27.44	37.04	46.56	F(2, 23) = 8.473, p < 0.002, motion binding and an SOA of 40 ms: *p* = 0.009, motion binding and an SOA of 300 ms: p = 0.031
2	-38.93	-24.77	-31.71	
3	9.21	17.01	25.13	
4	21.03	42.98	19.05	
5	-34.65	-15.07	-23.14	
6	9.19	23.12	13.92	
7	-51.26	-11.90	-4.11	
8	15.00	47.06	33.44	
9	17.04	33.06	37.51	
10	-9.02	17.60	44.61	
11	-21.98	-24.64	-13.83	
12	-9.15	-21.21	-17.99	

Table A-12. JNDs of Experiment 3 in Chapter 3.3. A repeated-measures analysis of variance (ANOVA) and multiple comparisons with Bonferroni correction are used to compare the JNDs in the motion binding condition and those in the non-motion binding condition. The table indicates a statistically significant difference between the motion binding and non-motion binding conditions.

Participant ID	JND			Repeated-measures analysis of variance (ANOVA) and Multiple comparisons with Bonferroni correction
	motion binding condition	non-motion binding condition (SOA:40ms)	non-motion binding condition (SOA:300ms)	
1	30.09	18.69	43.53	$F(2, 23) = 6.053$, $p < 0.008$, motion binding and an SOA of 40 ms: $p = 0.634$, motion binding and an SOA of 300 ms: $p = 0.002$
2	12.58	58.96	32.07	
3	33.57	17.39	47.07	
4	19.50	13.07	22.42	
5	9.24	24.54	25.06	
6	32.73	48.48	43.94	
7	35.97	45.94	80.11	
8	13.92	17.49	28.41	
9	21.99	20.25	28.21	
10	21.69	36.47	44.61	
11	38.91	43.54	68.33	
12	31.03	27.62	41.53	

Table A-13. Results of the face-to-face communication condition in Chapter 4.

Pair ID	Measurement Period (min:sec.msec)	Density of Synchronized Head Nods (nods / min)	Mean Phase Difference (ms)	SD (ms)	Kurtosis
1	04:42.85	8.7	110	260	0.6
2	10:01.17	4.7	250	430	-1.0
3	06:00.04	7.3	160	410	-0.6
4	05:31.63	7.1	110	410	-0.3
5	06:48.74	6.2	60	240	1.0
6	06:57.20	21	130	190	4.2
7	07:52.14	14	50	250	3.4
8	06:38.85	8.3	20	200	2.4
9	08:44.31	8.6	100	330	1.3
10	06:04.71	8.6	60	360	0.7
11	05:45.00	9.6	180	410	0.1
12	05:29.77	7.1	80	300	3.6

Table A-14. Results of the remote communication condition in Chapter 4.

Pair ID	Measurement Period (min:sec.msec)	Density of Synchronized Head Nods (nods / min)	Mean Phase Difference (ms)	SD (ms)	Kurtosis
1	05:29.83	4.5	90	470	-0.5
2	06:40.98	3.6	20	430	-0.1
3	07:19.59	3.0	140	440	-0.8
4	05:27.04	3.3	-50	580	-1.6
5	06:12.68	8.4	290	420	0.9
6	07:02.75	4.4	130	340	1.2
7	06:40.07	3.1	100	430	-0.4
8	05:55.76	6.6	-20	440	1.0
9	06:20.89	2.7	90	550	-1.0
10	06:35.28	6.7	110	310	0.9
11	07:11.30	4.0	-20	350	1.3
12	09:18.63	4.8	60	420	-0.2

Table A-15. Mean relative frequency of synchronized head nods for every 100 ms across all pairs (in the face-to-face communication and remote communication conditions)

Temporal Range	Face-to-face Communication Condition	Remote Communication Condition
— 900 ~ — 1000	0.025	0.024
— 800 ~ — 900	0.011	0.010
— 700 ~ — 800	0.014	0.018
— 600 ~ — 700	0.019	0.025
— 500 ~ — 600	0.019	0.023
— 400 ~ — 500	0.041	0.046
— 300 ~ — 400	0.030	0.029
— 200 ~ — 300	0.054	0.060
— 100 ~ — 200	0.082	0.081
0 ~ — 100	0.098	0.097
0 ~ 100	0.128	0.117
100 ~ 200	0.104	0.103
200 ~ 300	0.082	0.075
300 ~ 400	0.079	0.071
400 ~ 500	0.044	0.048
500 ~ 600	0.044	0.038
600 ~ 700	0.033	0.035
700 ~ 800	0.046	0.048
800 ~ 900	0.030	0.029
900 ~ 1000	0.019	0.022

Table A-16. The Delay time between the time on the stopwatch on the computer and that on the television screen. (ms)

	Computer	Television	Delay time
1	638	810	172
2	684	840	156
3	384	541	157
4	148	290	142
5	804	945	141
6	903	1059	156
7	461	601	140
8	553	725	172
9	64	220	156
10	122	294	172
11	555	727	172
12	683	866	183
13	802	958	156
14	265	422	157
15	404	560	156
16	621	761	140
17	697	853	156
18	820	992	172
19	53	224	171
20	330	502	172
21	766	922	156
22	904	1060	156
23	903	1059	156
24	120	291	171
25	305	461	156
26	523	662	139
27	754	894	140

28	924	1095	171
29	156	297	141
30	465	621	156
31	760	931	171
32	852	992	140
33	254	427	173
34	362	533	171
35	635	797	162
36	827	967	140
37	919	1090	171
38	198	354	156
39	553	694	141
40	267	423	156
41	359	530	171
42	732	903	171
43	228	400	172
44	554	694	140
45	499	655	156
46	167	323	156
47	167	461	171
48	630	802	172
49	829	1001	172
50	418	606	188
Mean			160
Standard Deviation			13