Studies on Human Mechanism of Audiovisual Temporal Integration by Event-related Potential

Jingjing Yang

March 2014

The Graduate School of Natural Science and Technology

(Doctor's Course)

OKAYAMA UNIVERSITY

Abstract

In everyday life, our brains integrate various kinds of information from different modalities to perceive our complex environment. Temporal proximity of multisensory stimuli is required for multisensory integration. We rarely pay attention to the differences in arrival time of auditory and visual inputs. In fact, the temporal factor of auditory and visual stimuli influences the processing of our brain. Many researches have shown that temporal asynchrony of visual-auditory stimuli can influence multisensory integration. But the neural activity of audiovisual temporal integration remains unclear. In present study, we used behavior and event-related potentials (ERPs) to examine the mechanism of audiovisual temporal integration.

First, the visual and auditory stimuli onset synchrony and only the visual stimulus were attended. This study used event-related potentials (ERPs) to demonstrate that onset synchronous task-irrelevant auditory stimuli affect the audiovisual integration. The behavioral results showed that the responses to audiovisual target stimuli were faster than that to unimodal visual target stimuli. Four ERP components related to audiovisual integration were observed. Those finding confirms the main neural activity of audiovisual integration when the visual and auditory stimuli onset synchrony.

Second, the visual and auditory stimuli onset asynchrony (SOA= ±400 ms, ±150 ms, 0 ms), only the visual stimulus was attended. Behavioral data and Event-Related Potentials (ERPs) were recorded. The behavioral results showed that the responses to temporal asynchronous AV stimuli were more accurate than unimodal visual stimuli. When the SOA was -150ms, the reaction time was the fastest and hit rate was the highest, The ERP results showed that the N1 latency of biomodal AV was earlier than the sum of unimodal auditory and unimodal visual stimuli in auditory preceding condition. These results suggested that the temporal asynchronous audiovisual stimuli enhanced the visual detection.

Finally, the differences between audiovisual integration elicited by reliable versus unreliable auditory stimuli were compared through behavioral and event-related potentials (ERPs) experiment, and found that the temporal reliability of auditory stimuli can modulate audiovisual integration at the early sensory processing stage.

Contents

Chapter 1

Introduction

Summary

This chapter describes the previous studies about audiovisual integration and the effect of temporal factor. Spatial and temporal proximity of multisensory stimuli is required for multisensory integration. Many researches have shown that temporal asynchrony of visual-auditory stimuli can influence multisensory integration. However, the neural mechanisms of asynchrony inputs were not well understood. Some researchers believe that humans have a relatively broad time window, in which stimuli from different modalities and asynchronous inputs tends to be integrated into a single unified percept. Others believe that the human brain can actively coordinate the auditory and visual input so that we do not notice the asynchronous inputs of multisensory stimuli.

1.1 Previous studies of audiovisual integration

Human brain is constantly deluged with various kinds of information from multiple sensory organs. When using a computer, people must look at the screen, listen to the sound from the speaker, touch the mouse or keyboard, and so on. Information from different sensory organs is often efficiently merged to form a unified and robust percept. This process is referred to as multisensory integration [1.1-1.3]. For example, the impact of a falling ball, the simultaneously generate multisensory information. In this event, the crashing ball not only reflects light to our eyes at the moment the ball strikes the ground but also creates air-borne vibrations and transmits them to our ear. Because of the different physical natures of these signals, neither has any effect on the other. Some of these physical signals can be transduced by the nervous system. The retina transduces light, and the cochlea transduces air-borne pressure waves into neural signals. The signals arrive at sensory-specific cortices to result in a distinct perception. Then, our neural system can automatically combine the neural signals from different sensory organs into a unified perception. Therefore, a ‘multisensory stimulus’ is actually, then, an event that generates several independent physical signals, each of which is simultaneously detectable by different types of sensory receptors [1.4].

A typical example of the audiovisual interaction is the McGurk effect, which was first described in a paper by McGurk and MacDonald in 1976[1.5]. When a video of one phoneme production is dubbed onto a sound recording of a different phoneme that is spoken, the perceived phoneme is a third, intermediate phoneme. For example, a visual /ga/ combined with an audio /ba/ is often heard as /da/. The McGurk effect demonstrates an interaction between hearing and vision in speech perception.

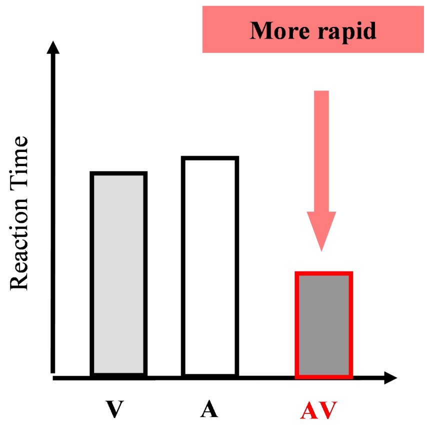

Fig.1.1 Average reaction time to unimodal visual stimuli unimodal auditory stimuli and multimodal audiovisual stimuli.

Such behavioral studies have shown, for example, that the simultaneous, or near-simultaneous, presentation of an auditory stimulus can influence the perceived temporal characteristics of a visual stimulus[1.6]. Electrophysiological studies in nonhuman primates and other mammals have shown that sensory cues from different modalities that appear at the same time and in the same location can increase the firing rate of multisensory cells in the superior colliculus[1.4,1.7]. Converging evidence from human behavioral research has demonstrated that stimuli from two or more sensory modalities presented in close spatial and temporal proximity can have a facilitative effect. Specifically, multimodal stimuli lead to faster detection times(Fig.1.1) and more accurate discrimination performance compared to unimodal stimuli. Spatially and temporally coincident audiovisual stimuli are detected more easily and more quickly than unimodal stimuli [1.8-1.10]. Human attention system enables us to focus on task-relevant information and to ignore that which is irrelevant [1.11]. However, task-irrelevant information cannot be completely ignored because humans must constantly monitor their environments.

Whether the auditory and visual stimuli are combined or not, it is relies on the stimulative accordance of temporal and spatial information [1.9]. In real-life, audiovisual stimuli onset asynchrony is ubiquitous phenomenon. When watching TV, We often see the mouth movement before hearing the sound. A flash of lightning, followed after some seconds by a rumble of thunder is the fact that sound travels significantly slower than light: 330 versus 300,000,000 meters per second.

Time is a special dimension of the environment in that there are no specific sense organs for its perception. Yet time can be appreciated in vision, hearing, and touch. Temporal disparity was easier in arousing the illusion if the visual stimulus was presented before the auditory stimulus by about 50-100ms [1.12]. "Due to the difference between the velocities of sound and light, auditory signals from an object located at a distance of about 15 m, reach our ears approximately 50 ms after light reaches the retina" [1.13]. Sugita ,Waka and Kopinskabelieve that the human brain can actively coordinate the auditory and visual input so that we do not notice the asynchronous inputs of multisensory stimuli[1.14,1.15]. They support the hypothesis of perceptual compensation. But Jorg,Derek and Stone refute this hypothesis[1.16-1.17]. They believe that humans have a relatively broad time window, in which stimuli from different modalities and asynchronous inputs tends to be integrated into a single unified percept.

1.2 The effect of temporal factor

1.2.1 Time window

Within the multisensory research community, the concept of a temporal window of integration has been well-described over 20 years ago. Why do we normally perceive the audiovisual events to be occurring simultaneously, even though the auditory and visual signals arrive at different time. Traditionally, some researchers believed that" humans had a relatively wide window for the integration of multisensory stimuli, and that we were therefore simply insensitive to small differences in the arrival time of signals to different sensory modalities" [1.18].

When the auditory and visual stimuli are not match in temporal characteristics, perceptual reports regarding the temporal characteristics of visual stimuli tend to be biased toward the temporal characteristics of the auditory stimuli [1.12,1.13,1.19]. For example, when participants were presented two flashes accompanied by three tone pips, they tended to report seeing three flashes rather than just two. Temporal disparity was easier in arousing the illusion if the visual stimulus was presented before the auditory stimulus by about 50-100ms. In the area of audiovisual speech perception, it has been observed that auditory speech has to lag behind matching visual speech, i.e., lip movements, by more than 250 ms for the asynchrony to be perceived [1.16,1.18,1.20]. Multisensory information falling within this window is highly likely to be integrated, whereas information falling outside is not [1.21].

1.2.2 Cognitive compensation

Sugita and Suzuki and Waka Fujisaki proposed that “auditory and visual inputs are coordinated not because the brain has a wide temporal window for auditory integration, as was previously thought, but because the brain actively changes the temporal location of the window depending on the distance of the visible sound source.” [1.14,1.15]. In the external environment, light travels far more rapidly than sound. Yet in human brain, the neural transmission rate is higher for auditory signals than for visual ones.

In experiment of Yoichi Sugita, the distance of light was 40 meters and sound was presented by headphone [1.14]. Observers judged whether the light was presented before or after the sound. It takes about 120 ms for sound to travel 40 m, and they found that the threshold for detecting the sound delay was 106 ms at a viewing distance of 40 m, so active compensation is likely to operate only for shorter distances than this. In perceiving the sound produced by the movement of a visible object, the brain coordinates the auditory and visual input so that no delay is noticed even though the sound arrives later. Coordination occurs because the brain uses information about distance that is supplied by the visual system to calibrate simultaneity.

In experment of Waka Fujisaki , a fixed audiovisual time lag for several minutes was presented before the formal experiment. Participents showed shifts in their subjective simultaneity responses toward that picture lag. Our brain can adjust for differences between the visual-auditory modalities in both phyysical and neural transmission time [1.15].

Recent research has uncovered two alternative means by which the brain ensures the continued perception of multisensory synchrony despite having to deal with asynchronous inputs. First is moveable window, the second is Temporal ventriloquism.

1.2.3 Moveable window

Sugita and Suzuki have suggested that" humans may have a 'moveable window' for multisensory integration, rather than having a stationary temporal window for multisensory integration"[1.14]. In a Psychophysical study, Sugita and Suzuki believed that "the window for multisensory temporal integration actually moves as audiovisual stimuli become more distant from us, in order to accommodate the fact that sound will increasingly lag behind vision with increasing distance" [1.18]. The existence of a moveable window for multisensory simultaneity, to accommodate the typical delay in auditory arrival times, is consistent with previous studies showing that people (and animals) are more sensitive to audiovisual asynchrony when sound leads vision than vice versa. However, Sugita and Suzuki found that this moveable window only operates up to a distance of around 10 metres, but not with larger distance, hence explaining why our perception of synchrony breaks down for more distant audiovisual events [1.14].

1.2.4 Temporal ventriloquism

Ventriloquism is most commonly demonstrated in the shift of sound toward the location of a visual event. Numerous studies have shown that vision has a strong impact on spatial information processing in audition[1.13].This impact is most clearly demonstrated in the spatial ventriloquist effect. The spatial ventriloquist effect is a very robust phenomenon. For example, when we are watching TV, we believe that the sound from people's mouth and not from the speaker of TV. The ventriloquism effect suggests that the visual system is dominant over the auditory system when it comes to spatial localization.

Temporal ventriloquism shows that a sound presented in close temporal proximity to a visual stimulus can alter the perceived temporal dimensions of the visual stimulus (temporal ventriloquism or binding) [1.17]. Morein-Zamir and colleagues have shown that the perceived time of arrival of a visual event can be ventriloquized into temporal alignment with a subsequently presented sound[1.22] . “The phenomenon of temporal ventriloquism not only corrects for differences in the time of arrival of stimuli from different sensory modalities, but is also involved in the synchronization of the rate at which sensory events are perceived to occur” [1.14].

In Morein-Zamir's experiment, the subjects were asked to perform a temporal order judgment task on the onsets of two LEDs. In this experiment a sound was presented before the first onset and after the second onset, compared to a neutral condition in which the sound coincided with the LED onsets. It seemed as if the visual onset was pulled in time towards the auditory onsets, which made temporal order judgment of the visual events easier. Ventriloquism and temporal ventriloquism show that one modality can bias another modality in the spatial and temporal domain. These effects suggest that the auditory modality is dominant in the temporal domain and the visual modality is dominant in the spatial domain [1.22].

1.3 The purpose of the present dissertation

The neural mechanism of audiovisual integration is very complex. Previous studies have shown that an audiovisual stimulus with spatial and temporal coincidence can generate the largest integration effects. The main aim of this research was to investigate the brain activities of audiovisual temporal integration using behavioral and electroencephalography (EEG) with high temporal resolution and to elucidate the mechanism of audiovisual temporal integration in humans. First, the effects of temporal synchrony on audiovisual integration were investigated. Second, the effects of the temporal asynchrony on audiovisual integration were investigated. Finally, the differences between audiovisual integration elicited by reliable versus unreliable auditory stimuli were compared.

1.4 The contents of the dissertation

The dissertation is composed of six chapters. The first and second chapters instructed the previous studies and methods of our study. The third, fourth and fifth chapters investigated the temporal characteristics of stimuli in human audiovisual integration.

Chapter 1 describes the previous studies about audiovisual integration and the effect of temporal factor. Previous studies have shown that an audiovisual stimulus with spatial and temporal coincidence can generate the largest integration effects.

Chapter 2 describes the concept of Electroencephalogram (EEG) and Event-related potentials (ERPs), and illustrates the method of Event-related potentials analysis.

Chapter 3 describes the first experiment, the visual and auditory stimuli onset synchrony and only the visual stimulus was attended. This experiment used event-related potentials (ERPs) to demonstrate that onset synchronous task-irrelevant auditory stimuli affect the audiovisual integration.

Chapter 4 introduces the second experiment. The effects of the temporal characteristics on audiovisual integration in a visual attention task were investigated. The visual and auditory stimuli onset asynchrony (SOA= ±400 ms, ±150 ms, 0 ms), only the visual stimulus was attended. Behavioral data and Event-Related Potentials (ERPs) were recorded.

Chapter 5 compared the differences between audiovisual integration elicited by reliable versus unreliable auditory stimuli through behavioral and event-related potentials (ERPs) experiment.

Chapter 6 provides a general conclusion based on the findings of this study and future challenges.

References

[1.1] Lovelace CT, Stein BE, Wallace MT. An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. Brain Res Cogn Brain Res 2003; 17:447–453.

[1.2] Frassinetti F, Bolognini N, Ladavas E. Enhancement of visual perception by crossmodal visuo-auditory interaction. Exp Brain Res 2002;147:332–343.

[1.3] Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. Nature Reviews Neuroscience, 9(4), 255−266.

[1.4] Meredith, M.A. On the neuronal basis for multisensory convergence: A brief overview. Cogn. Brain Res. 2002, 14, 31–40.

[1.5] McGurk H, MacDonald J. Hearing lips and seeing voices. Nature 1976; 264: 746-748.

[1.6] Lippert M, Logothetis NK, Kayser C. Improvement of visual contrast detection by a simultaneous sound. Brain Res 2007; 1173:102–109.

[1.7] Wallace, M. T. (1998). Multisensory integration in the superior colliculus of the alert cat. Journal of Neurophysics, 80, 1006 – 1010.

[1.8] Posner MI. Orienting of attention. Q J Exp Psychol 1980; 32:3–25.

[1.9] D. Talsma, M.G. Woldorff. Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. J Cogn Neurosci., 17, 2005, pp. 1098-1114.

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

[1.11] Carrasco M, Ling S, Read S. Attention alters appearance. Nat Neurosci. 2004; 7:308–313

[1.12] Slutsky DA, Recanzone GH. Temporal and spatial dependency of the ventriloquism effect. Neuroreport 2001; 12: pp. 7–10.

[1.13] D.Talsma, D.Senkowski, M.G..Woldorff. Intermodal attention affects the processing of the temporal alignment of audiovisual stimuli. Exp Brain Res. 2009; 198: pp. 313–328.

[1.14] Sugita, Y., & Suzuki, Y. (2003). Audiovisual perception: Implicit estimation of sound-arrival time.Nature, 421, 911.

[1.15] Waka Fujisaki,Shinsuke Shimojo, Makio Kashino and Shinya Nishida. Recalibration of audivisual simultaneity. Nature neuroscience,2004;7:pp773-778

[1.16] Lewald, J., & Guski, R. (2004). Auditory-visual temporal integration as a function of viewing distance: No compensation for sound-transmission time in human perception. Neuroscience Letters, 357, 119 – 122.

[1.17] Stone, J.V., Hunkin, N.M.,Porrill, J.,Wood,R.,Keeler,V.,Beanland,M.,etal.Whenisnow? Perception of simultaneity.(2002). Proceedings of the Royal Society of London Series B: BiologicalScience;p268.

[1.18] Spence, C., & Squire, S. (2003). Multisensory integration: Maintaining dispatch the perception of synchrony. Current Biology, 13, 519 – 521

[1.19] Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. Brain Research. Cognitive Brain Research, 14, 147 – 152.

[1.20] Jones, J. A., & Jarick, M. (2006). Multisensory integration of speech signals: The relationship between space and time. Experimental Brain Research, 174, 588–594.

[1.21] Hans Colonius1 and Adele Diederich. The optimal time window of visual–auditory integration: a reaction time analysis. Integrative Neuroscience.2010;4:pp.1-8.

[1.22] Morein-Zamir S, Soto-Faraco S, Kingstone A (2003). Auditory capture of vision: examining temporal ventriloquism. Brain Res Cogn Brain Res 17:154 – 163.

Chapter 2

Electroencephalogram and Event-related potentials

Summary

The aim in this chapter is to describe the concept of Electroencephalogram(EEG) and Event-related potentials(ERPs), and to illustrate the method of Event-related potentials analysis. Electroencephalography is the recording of electrical activity along the scalp produced by the firing of neurons within the brain. An event-related potential is the measured brain response that is the direct result of a specific sensory, cognitive, or motor event.

2.1 Electroencephalogram (EEG)

Electroencephalography (EEG) is the recording of electrical activity along the scalp produced by the firing of neurons within the brain [2.1]. The brain's electrical charge is maintained by billions of neurons. The recordings show fluctuations with time that are often rhythmic in the sense that they alternate regularly When the neurons of the human brain process information, they do so by changing the flow of electrical currents across their membranes. These changing currents generate electric and magnetic fields that can be recorded from the surface of the scalp. The electric fields are measured by attaching small electrodes to the scalp.

Fig.2.1 Scene of EEG experiment

The potentials between different electrodes are then amplified and recorded as the electroencephalogram(EEG), which means the writing out of the electrical activity of the brain (that which is inside the head). The human EEG was first recorded in 1924 by a German psychiatrist named Hans Berger [2.2]. The electric potential generated by an individual neuron is far too small to be picked up by EEG. EEG activity therefore always reflects the summation of the synchronous activity of thousands or millions of neurons that have similar spatial orientation.

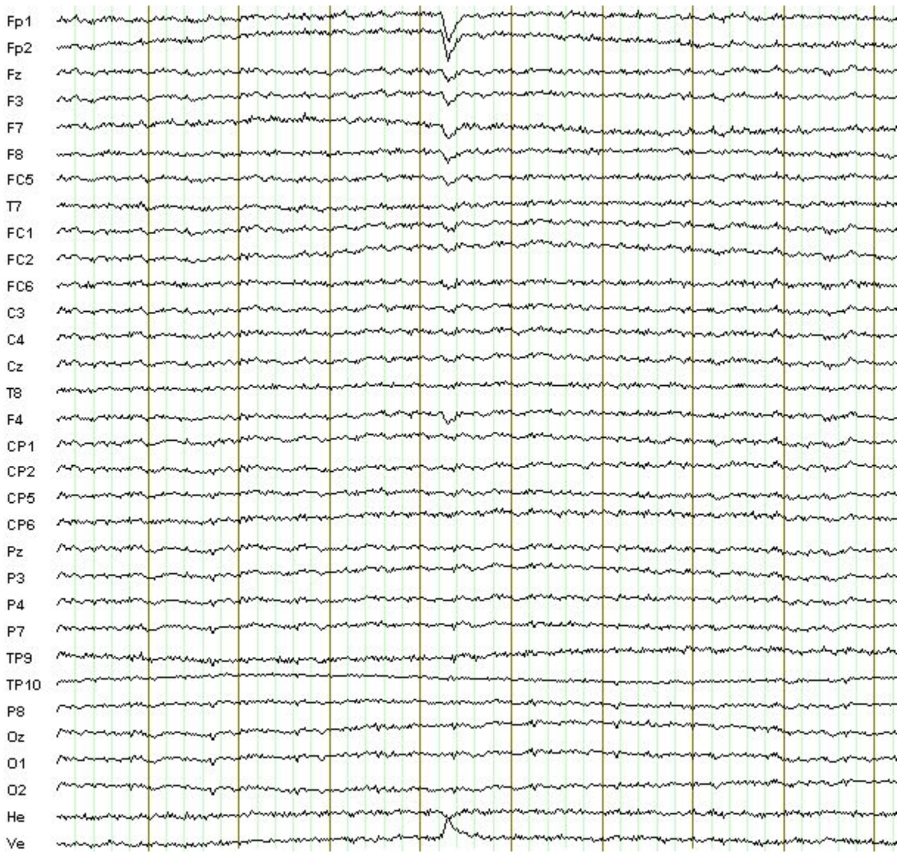

Fig.2.2 Raw data of EEG from 32 electrodes

EEG is non-invasive and does not involve any X-rays, radiation, or injections. EEGs have been used for many years and are considered very safe. In addition, EEGs have a high temporal resolution compared with techniques and are capable of detecting changes in electrical activity in the brain on a time scale in the millisecond region. Figure 2.1 shows the scene of an EEG experiment, and Figure 2.2 shows recorded EEG data. EEG has several limitations. Most important is its poor spatial resolution. EEG is most sensitive to a particular set of post-synaptic potentials: those generated in superficial layers of the cortex, on the crests of gyri directly abutting the skull and radial to the skull. Dendrites, which are deeper in the cortex, inside sulci, in midline or deep structures (such as the cingulate gyrus or hippocampus), or producing currents that are tangential to the skull, have far less contribution to the EEG signal. EEG recordings do not directly capture axonal action potentials.

The EEG is typically described in terms of rhythmic activity and transients. The rhythmic activity is divided into bands by frequency. To some degree, these frequency bands are a matter of nomenclature (i.e., any rhythmic activity between 8–12 Hz can be described as "alpha"), but these designations arose because rhythmic activity within a certain frequency range was noted to have a certain distribution over the scalp or a certain biological significance. Frequency bands are usually extracted using spectral methods (for instance Welch) as implemented for instance in freely available EEG software such as EEGLAB. Most of the cerebral signal observed in the scalp EEG falls in the range of 1–20 Hz.

2.2 Event-related potentials (ERPs)

The EEG proved to be a useful source in recording brain activity over the ensuing decades. However, it tended to be very difficult to assess the highly specific neural process that are the focus of cognitive neuroscience because using pure EEG data made it difficult to isolate individual neurocognitive processes. Event-related potentials (ERPs) offered a more sophisticated method of extracting more specific sensory, cognitive, and motor events by using simple averaging techniques. Event-related potentials (ERPs) refer to averaged EEG responses that are time-locked to more complex processing of stimuli. This technique is used in cognitive science, cognitive psychology, and psychophysiological research. An ERP is the measured brain response that is the direct result of a specific sensory, cognitive, or motor event. More formally, it is any stereotyped electrophysiological response to a stimulus. The evoked potential is the change in brain's electrical activity that follows a stimulus such as a click. The evoked potential is usually too small to recognize amidst all of the other activity recorded from the brain. When one hears a click, the brain may also be evaluating what eyes are seeing, thinking about what to have for supper, or controlling my hand as it is writing. Each of these activities will be associated with its own particular electric and magnetic fields. The specific patterns associated with the click can be measured if the click is repeated many times. Each time the click occurs, the evoked potential to the click will be the same, but all of the other activity occurring in the brain will change. If we average all of the recordings together, the other activities, which are different every time the click occurs, tend to cancel each other out, leaving the response to the click.

The averaging technique can be applied to EEG signals that occur before a behavioral response, such as pressing a button. This allows us to examine the processes in the brain that lead to the action. The average potentials that occur before, during or after a timing-cue are called event-related potentials.

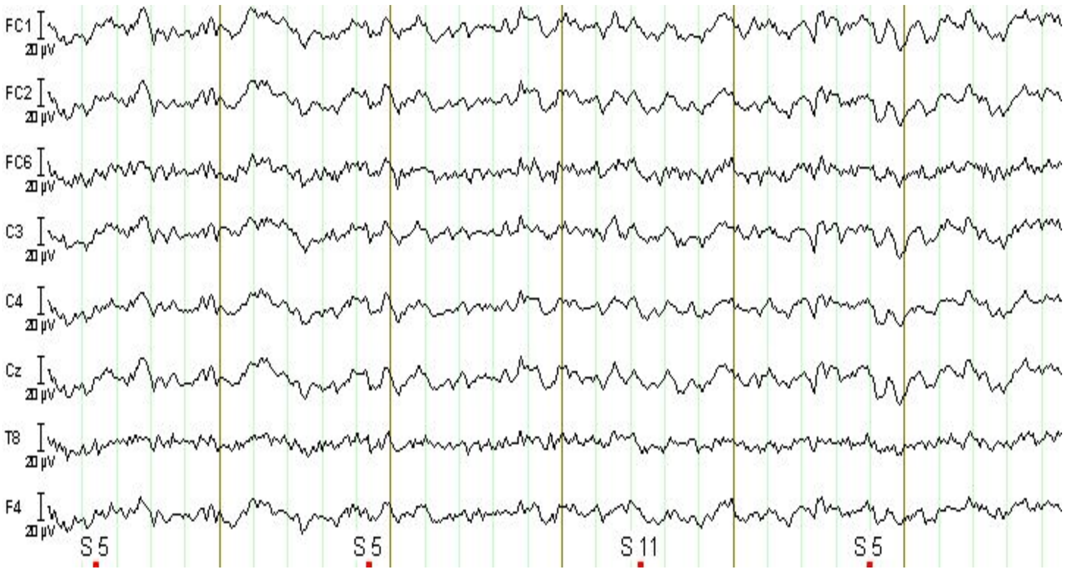

Fig.2.3 Raw data of EEG. The red point shown the list separator, in this time the stimulus of S5 was started to present.

1) Raw data of EEG

The EEG data was recorded from the experiment start to the end, only the event related potentials were subservient. The separators were listed when stimulus was presented such as S5, S11 (Fig.2.3).

2) Segmentating the raw data

The ERP data were event related and were segmented from EEG data through the list separator. As shown in Fig.2.4, The ERP data of one stimulation type were divided.

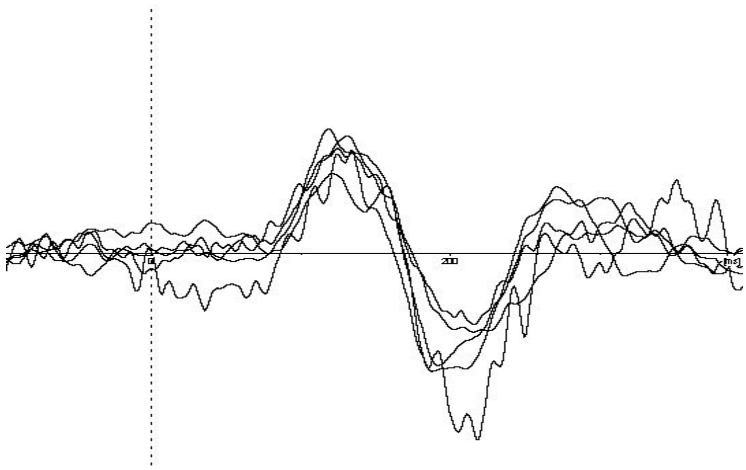

Fig.2.4 EEG signals were divided into epochs from 100ms before stimulus onset to 400ms after stimulus onset.

3) Averaging ERP data

In the end the ERP data of one stimulation type about 200 times were averaged. Then the real signals of ERP were emerged and the noises were taken out(Fig.2.5).

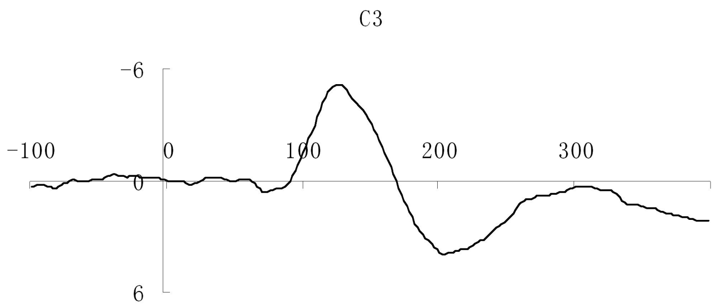

Fig.2.5 The averaged ERP data from C3 electrode.

ERP waveforms consist of a series of positive and negative voltage deflections, which are related to a set of underlying components. Though some ERP components are referred to with acronyms (e.g., contingent negative variation – CNV, error-related negativity – ERN, early left anterior negativity – ELAN, closure positive shift – CPS), most components are referred to by a letter (N/P) indicating polarity (negative/positive), followed by a number indicating either the latency in milliseconds or the component's ordinal position in the waveform. For instance, a negative-going peak that is the first substantial peak in the waveform and often occurs about 100 milliseconds after a stimulus is presented is often called the N100 (indicating its latency is 100 ms after the stimulus and that it is negative) or N1 (indicating that it is the first peak and is negative); it is often followed by a positive peak, usually called the P200 or P2. The stated latencies for ERP components are often quite variable. For example, the P300 component may exhibit a peak anywhere between 250ms – 700ms.

2.3 Method of ERP analysis in the audiovisual integration study

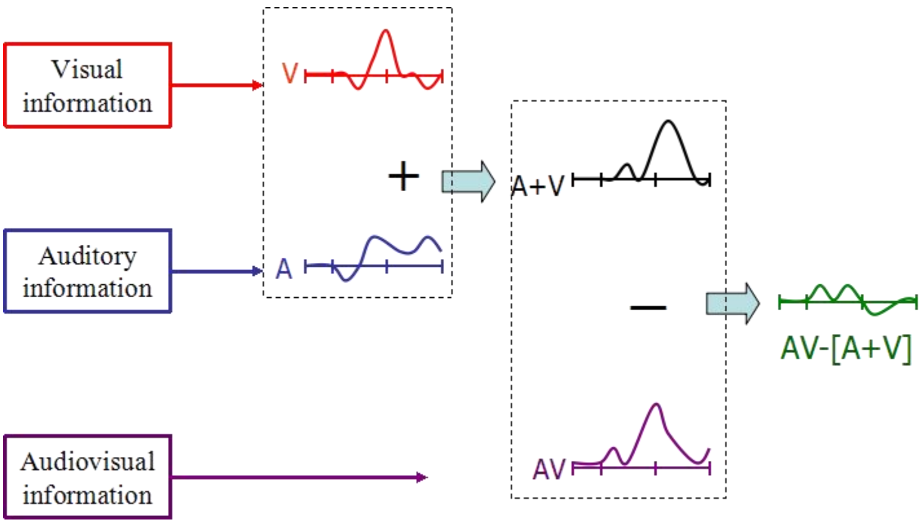

Fig.2.6 Method of ERP analysis in the audiovisual integration study

Recent ERP studies quantified auditory-visual interactions by the difference wave [AV- (A + V)] obtained by subtracting the sum of the responses to the unimodal stimuli from the response to multimodal stimuli(Fig.2.6). The method was first developed in animal single-cell recordings [2.3], and it was then adapted for use in human ERP studies [2.4].

We assumed the neural activities induced by the multimodal (AV) stimulus were equal to the sum of the neural activities induced separately by the auditory (A) and the visual (V) stimulus, in addition to the putative neuronal activities induced uniquely by multimodal stimulation (auditory-visual interactions). This assumption is valid only while the stimulus analysis is not "contaminated" by late activities related to target processing (P3 waves) or by activities related to the response selection or motor processes (all of these activities are common to all three stimulus types A, V, and AV). We may therefore use the summative model to estimate the AV interactions:

$$ERP(AV)=ERP(A)+ERP(V)+ERP(AVinteraction)$$

This expression is valid regardless of the nature or configuration of the intracerebral generators and is based on the law of superposition of electric fields [2.5].

References

[2.1] Niedermeyer E. and da Silva F.L. Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Lippincot Williams & Wilkins. (2004).

[2.2] Tatum, W. O., Husain, A. M., Benbadis, S. R. "Handbook of EEG Interpretation" Demos Medical Publishing (2008)

[2.3] Stein BE, Meredith MA. The Merging of the Senses. Cambridge: MIT Press;(1993), 123-156.

[2.4] Giard MH, Peronnet F. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci*(1999), 11: 473-490.

[2.5] Jennifer L., Mozolic, Christina E., Hugenschmidt, et al. Modality-specific selective attention attenuates multisensory integration. Exp Brain Res (2008), 184: 39-52.

Chapter 3

The effects of temporal synchrony on audiovisual integration

Summary

This chapter demonstrated that onset synchronous task-irrelevant auditory stimuli affect the audiovisual integration. The behavioral results showed that the responses to audiovisual target stimuli were faster than that to unimodal visual target stimuli. Moreover, the ERPs were recorded in response to unimodal auditory (A), unimodal visual (V) and bimodal (AV) stimuli. Cross-modal interactions were estimated using the additive [AV - (A + V)] model. Four ERP components related to audiovisual integration were observed: (1) over ce ntral and occipital areas at around100 to 160ms; (2) over the central and occipital areas at around 160 to 200ms; (3) over the occipital areas at around 200 to 240ms. (4) over frontal-central areas at around 280to320ms. These findings confirmed the main neural activity of audiovisual integration. In addition, our study provided evidence that multimodal integration can be generated even if the auditory stimulus was task- irrelevant.

3.1 Background

Human brain is capable of combining information from multiple sensory modalities. It has been well established that bimodal audiovisual stimuli are detected and discriminated more accurately than either visual or auditory unimodal stimuli presented alone [1.1,1.2]. Whether the auditory and visual stimuli are combined or not, it is relies on the stimulative accordance of temporal and spatial [1.6,1.8,1.9].

Such behavioral studies have shown, for example, that the simultaneous, or near-simultaneous, presentation of an auditory stimulus can influence the perceived temporal characteristics of a visual stimulus [1.2,1.6]. Electrophysiological studies in nonhuman primates and other mammals have shown that sensory cues from different modalities that appear at the same time and in the same location can increase the firing rate of multisensory cells in the superior colliculus[3.1,1.4]. Spatially and temporally coincident audiovisual stimuli are detected more easily and more quickly than unimodal stimuli [1.6,1.9]. Human attention system enables us to focus on task-relevant information and to ignore that which is irrelevant [1.8,1.11]. However, task-irrelevant information cannot be completely ignored because humans must constantly monitor their environments. An informative sound can reduce the uncertainty about the timing of the visual display, and significantly improve detection rates. A completely redundant sound cannot alter detection rates, but it can speed reaction times [3.2].

The audiovisual integration elicited by task-irrelevant sound has been investigated through behavioral experiment, but when and where the integration is accomplished in the brain is not well understood. In addition, the previous studies mainly limited their analysis of ERP to the first 200ms after stimulus presentation. Few assessed the brain activity after 200ms following presentation of the stimulus. Present study designed a visual attention task that included centrally presented audiovisual stimuli, and ascertained the effect of task-irrelevant auditory stimuli on multimodal audiovisual integration.

3.2 Methods

3.2.1 Subjects

Thirteen healthy volunteers (23 to 31 years, mean: 25.9 years, 5 females) from the Okayama University participated in the experiment. All participants were right hand and had normal or corrected-to-normal vision and normal hearing. After receiving a full explanation of the purpose and risks of the research, subjects provided written informed consent for all studies as per the protocol approved by the institutional research review board. The participants provided written informed consent for their participation in this study, which was previously approved by the ethics committee of Okayama University.

3.2.2 Stimuli

Unimodal visual (V) stimuli consisted of vertical (0°,standard stimulus) and horizontal (90°, target stimulus) Gabor gratings (2°diameter, 2 cycles/degree), which could appear 4°below the fixation point (duration of 23ms).

Unimodal auditory(A) stimuli consisted of 1000 Hz sinusoidal tone, presented at a sound pressure level of 70 dB with two speakers that were placed on either side of the display (duration of 40 ms ,including linear rise and fall times of 5 ms).

The bimodal AV stimulus was always composed of the simultaneous presentation of both unimodal A and V stimuli. The auditory (A) stimuli provided reliable and fixed temporal information.

3.2.3 Task and procedure

Experiments were conducted in a dimly lit, sound-attenuated and electrically shielded room(laboratory room, Okayama University, Japan). Visual stimuli were presented on a 17-in display. The background luminance was set to 2.7cd/m^2 and was perceived as neutral grey. Participants viewed the display from a distance of 57 cm, with their head placed on a custom-made chin rest.

1) Preparation Experiment

In order to determine one intermediate contrast level so that the accuracy in detecting the stimuli was over 80%, a pre-experiment were conducted to all experiments. In pre-experiment, only visual stimuli were presented. The visual (V) stimuli consisted of vertical (standard stimulus) and horizontal (target stimulus) Gabor gratings (2°diameter, 2 cycles/degree), which could appear 4°below the fixation point (duration of 23ms)(Fig.3.1). The contrast levels were 2%, 2.5%, 3%, 5% and 7% in units of Michelson contrast [(max−min)/(max+min)], with max and min being the maximal and minimal values of the Gabor patch(Fig.3.2).

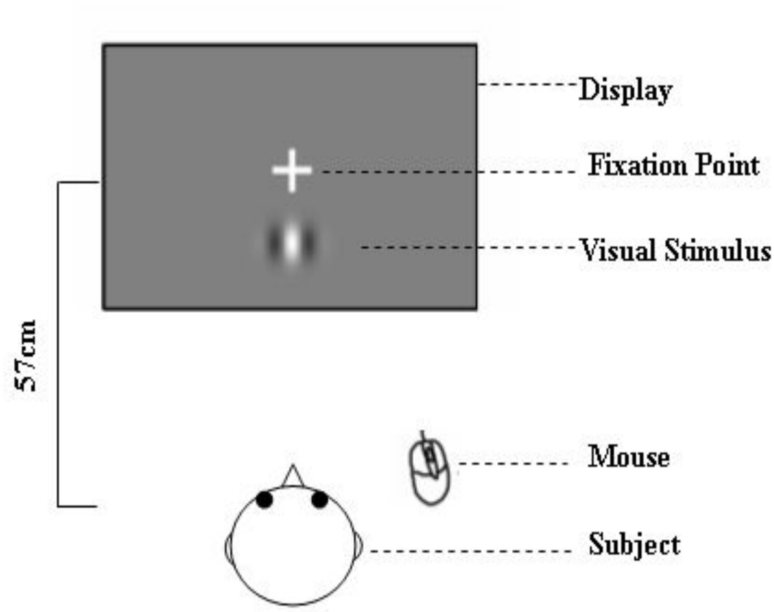

Fig.3.1 The distance is 57cm from subject to display.The Visual stimulus was presented 4° below the plus. The subject's task was to detect the target stimuli as quickly and accurately as possible and to respond by pressing the left button of a computer mouse.

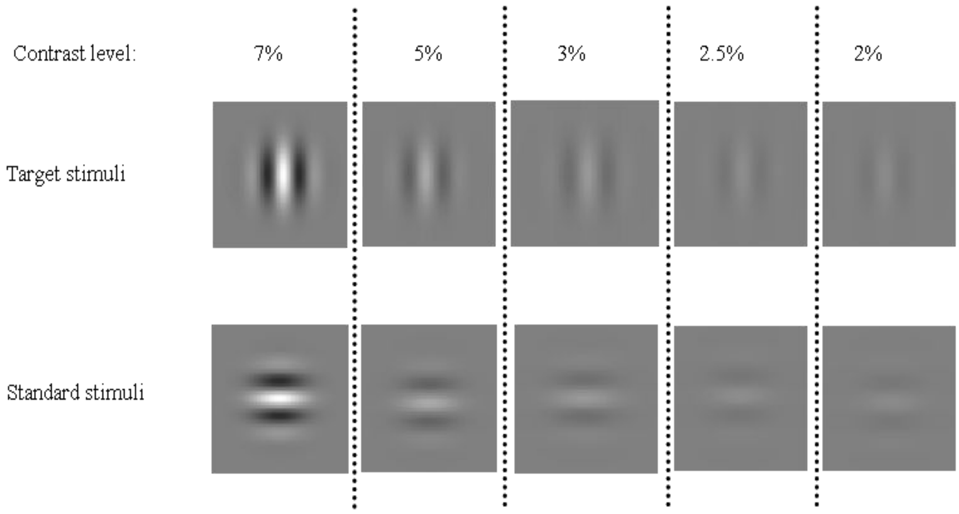

Fig.3.2 The standard and target stimuli of preparation experiment. The contrast of these ranged from 2 to 7% in units of Michelson contrast [(max−min)/(max+min)], with max and min being the maximal and minimal values of the Gabor patch.

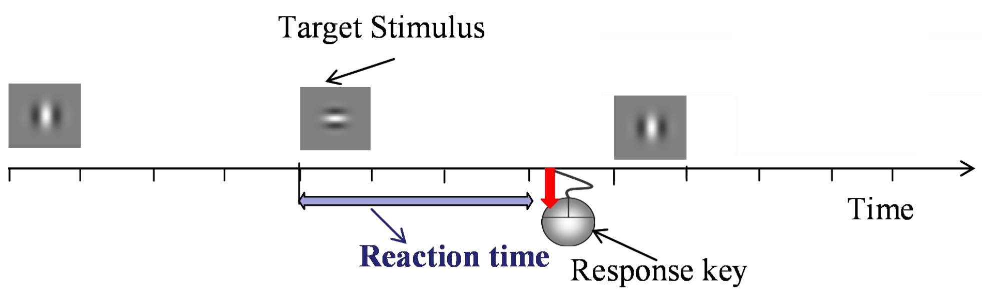

Fig.3.3 Recording method of reaction times

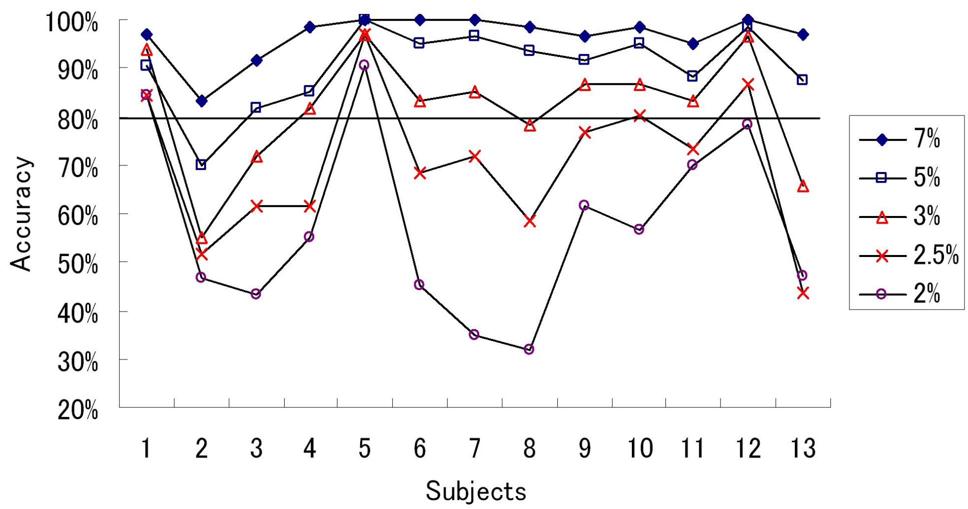

Fig.3.4 The averaged accuracies of 5 contrast levels from 13 subjects. According to the subject one kind of V contrast level which correct detection above 80% was selected. For example, when the contrast level was 5% the accuracy of subject 3 aboved 80%.

The pre-experiment had 3 sessions. Each session consisted of 200 visual stimuli. The frequency of target stimulus was 50% in each stimulus type. As shown in Fig. 3.1, the distance is 57cm from subject to display, and the inter stimulus interval (ISI) varied randomly between 600 and 1000ms. The experiment was conducted in a dimly lit, sound-attenuated room. Throughout the experiment, the stimuli were randomly presented. And the subjects were required to fix their eyes on the centrally presented fixation point on the display and to detect all target stimuli as quickly and accurately as possible and responding by pressing the left button of a computer mouse with their right hands. Response time data were analyzed for correct responses to target stimuli. Recording method of reaction times is shown in Fig.3.3. The results of averaged accuracies were showed in Fig.3.4.

2) Audiovisual Integration Experiment

Audiovisual integration experiment contained three stimulus types which are unimodal visual (V) stimuli, unimodal auditory (A) stimuli and bimodal audiovisual (AV) stimuli. According to the subject one kind of V contrast level was selected from preparation experiment, which the correct detection was above 80%, as shown in Fig.3.4. The bimodal AV stimulus was always composed of the simultaneous presentation of both unimodal A and V stimuli.

The V and AV stimuli contained standard and target stimulus. The AV standard or target stimulus consisted of the simultaneous presentation of both unimodal A and V standard or target stimuli. The participants' task was to attend only to the visual modality and respond only to the V and AV target stimuli. For each type of stimulus, target stimuli were 20%. The inter stimulus interval (ISI) varied randomly between 800 and 1400ms(Fig.3.5). Throughout the experiment, the stimuli were randomly presented. And the subjects were required to fix their eyes on the centrally presented fixation point on the display and to detect all target stimuli as quickly and accurately as possible and responding by pressing the left button of a computer mouse with their right hands.

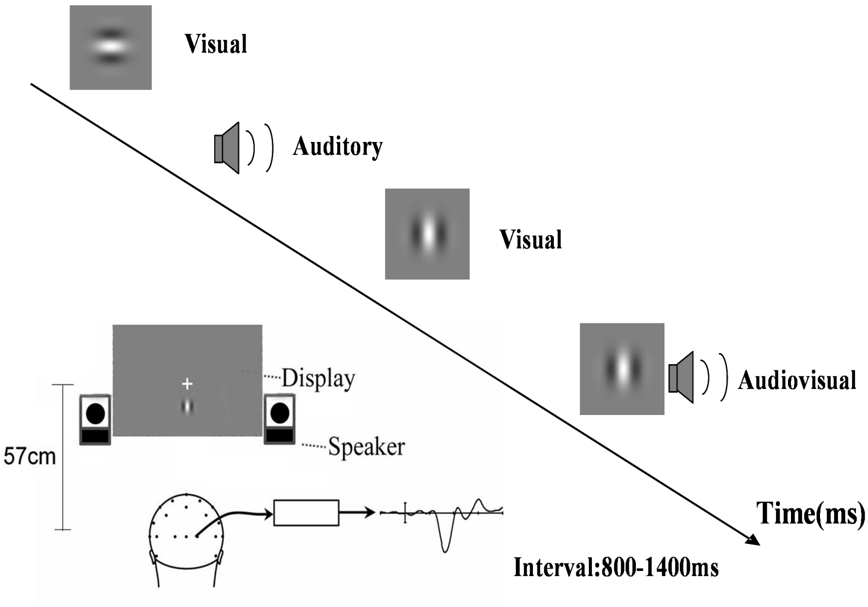

Fig.3.5 Experimental arrangement of the display and speakers. The V stimulus (Gabor gratings) was presented on a display at an angle of 4°below the fixation point. The A stimulus was presented through two speakers placed on either side of the display. Participants viewed the display from a distance of 57 cm. The Auditory (A), Visual (V) and 5 types Audiovisual (AV) stimuli were presented with equal probability. The participants only attend to the V stimuli and only respond to the V target stimuli.

3.2.4 Data Analysis

1) Behavioral analysis

Stimulus presentation was controlled by a personal computer running "Presentation" software (Neurobehavioral Systems, Albany, CA). Hit rates were the number of correct responses to target stimuli divided by the total number of target stimuli. Response time data were analyzed for correct responses to target stimuli. Reaction times (RTs) and hit rates (HRs) were computed separately for each stimulus type and were analyzed using paired sample t-texts.

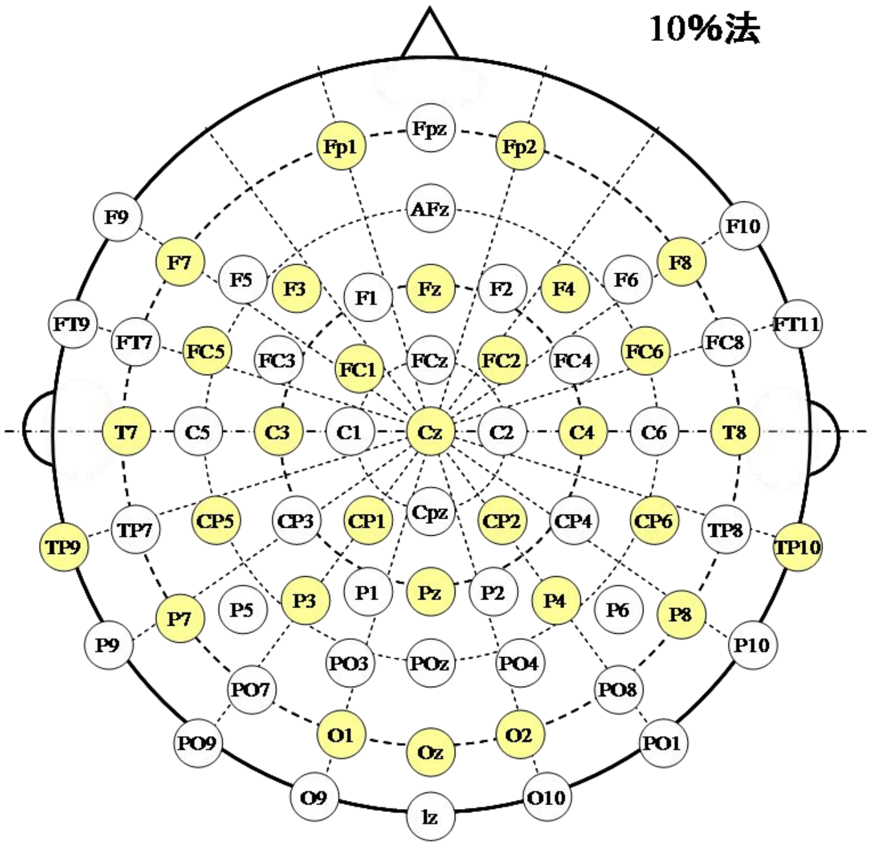

Fig.3.6 Pointing places of electrodes with 10-20% international standard system

2) Event-related potential (ERP) analysis

Electrophysiological signals were recorded using BrainAmp MR plus (Germany) via 32 electrodes mounted on an electrode cap(Easy-cap, Herrsching Breitbrunn, Germany) according to the 10/20 international standard system(Fp1, Fp2, Fz, F3, F4, F7, F8, FC1, FC2, FC5, FC6, Cz, C3, C4, CP1, CP2, CP5, CP6, Pz, P3, P4, P7, P8, TP9, TP10, T7, T8, O1, O2, Oz)(Fig.3.6). All signals were referenced to the linked ears electrode. All electrode impedance was maintained below 5kΩ. Horizontal eye movements were recorded from the outer canthus of the left eye; eye blinks and vertical eye movements were recorded from an electrode below the left eye. The ERP analysis was carried out using Brain Vision Analyzer software (version 1.05, Brain Products GmbH, Munich, Bavaria, Germany). Raw signals were digitized with a sample frequency of 500Hz with a 60Hz notch filter and all data were stored digitally for off-line analysis. An artifact criterion over ± 80μV was used at all channels to reject trials with noise transients. The data were then averaged for each stimulus type following digital filtering using a band-pass filter of .01–30Hz. After filtering, final grand-averaged data were obtained across all subjects for each stimulus type. Only the ERPs elicited by standard stimuli were analyzed. Continuous EEG signals were divided into epochs from 100ms before stimulus onset to 400ms after stimulus onset.3.3 Results

3.3.1 Behavioral results

In behavioural experiment, average reaction times and response accuracy across subjects in behavioural experiments were shown in Fig.3.7 Reaction times (RTs) and hit rates (HRs) were computed separately for each stimulus type and were analyzed using Paired sample t-texts. HRs to AV target stimuli were more accurate than unimodal V [t(12)=-2.289,p=.041]. RTs of AV stimuli were significantly faster than unimodal V [t(12)=3.142,P=.009].

The ERPs elicited by the standard stimuli were analyzed. The EEG and EOG signals were amplified and band-pass filtered with an analog filter of 0.01 - 100 Hz at a sampling rate of 500 Hz. EEG and EOG signals were divided into epochs from 100 ms before the stimulus onset to 600 ms after onset, and baseline corrections were made against -100-0ms.

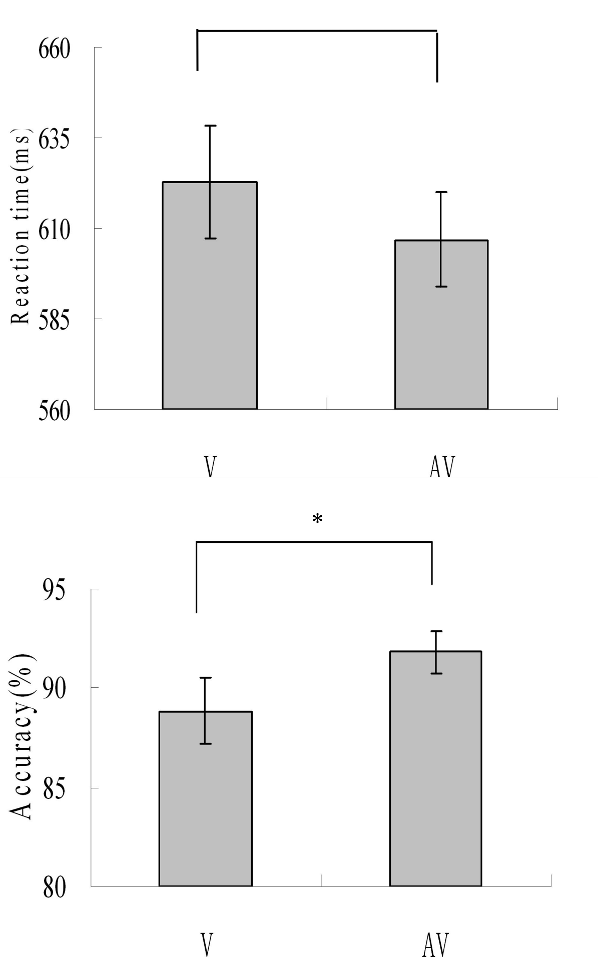

Fig.3.7 Average reaction time and response accuracy in behavioral experiment (* P<0.01, ** p < 0.001).

3.3.2 ERP results

Fig3.8 and Fig.3.9 shows the grand-averaged ERPs elicited by unimodal A and unimodal V standard stimuli. The auditory ERPs included three typical waves: (1) the negative N1 wave peaked around 100ms post-stimulus at fronto-central sites (126.9±7.3ms at Cz)(2)the positive P2 wave peaked around 200ms post-stimulus (214.8±12.6ms at Cz); (3) the negative N250 wave peaked (320.5±16.6ms at Cz). The visual ERPs presented a negative N1wave (peaking at around 222ms at Pz) followed by a positive P1wave (peaking at around 280ms at Pz).The audiovisual ERPs components were similar to the auditory ERPs.

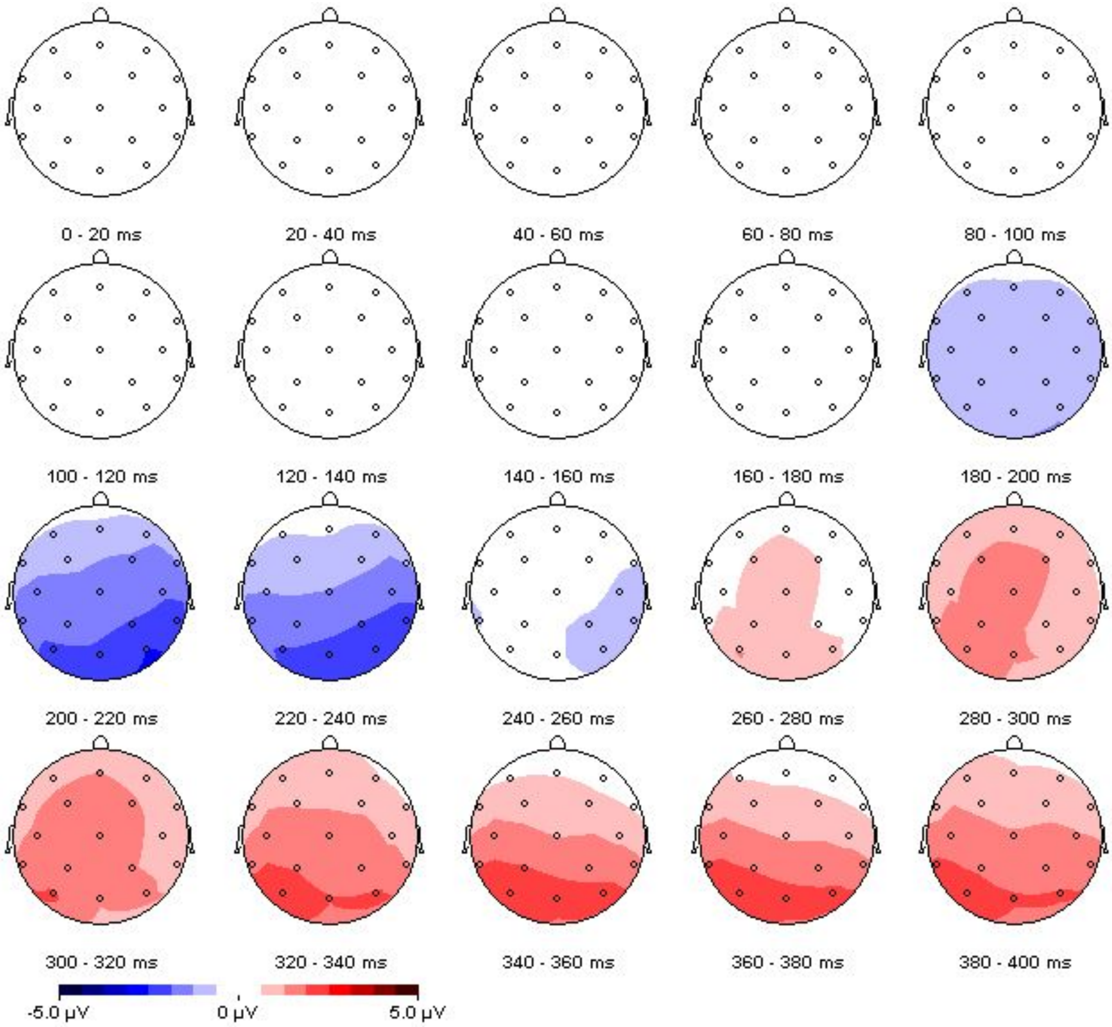

Fig.3.8 The scalp topographies of grand-averaged event-related potentials elicited by unimodal visual (V)stimuli

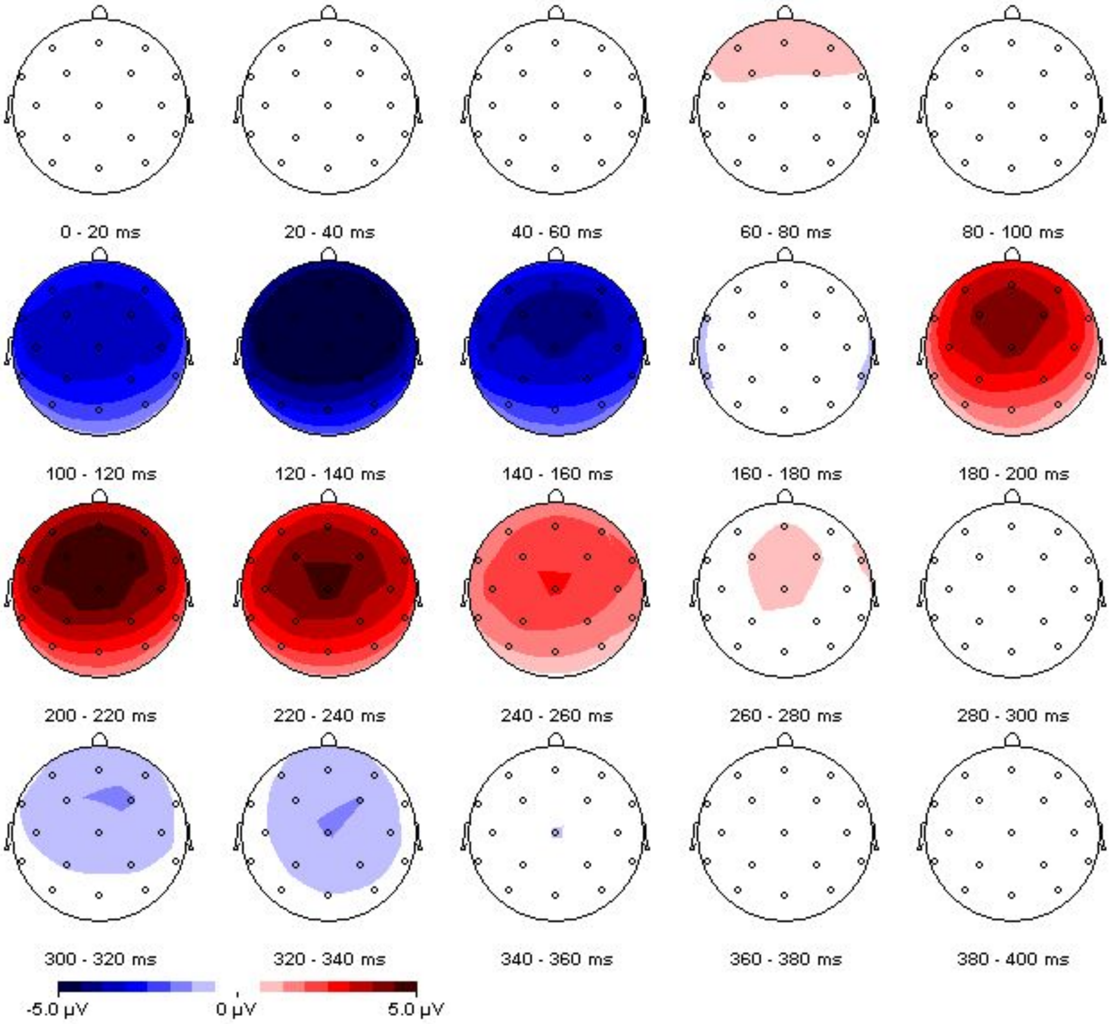

Fig.3.9 The scalp topographies of grand-averaged event-related potentials elicited by unimodal auditory (A)stimuli

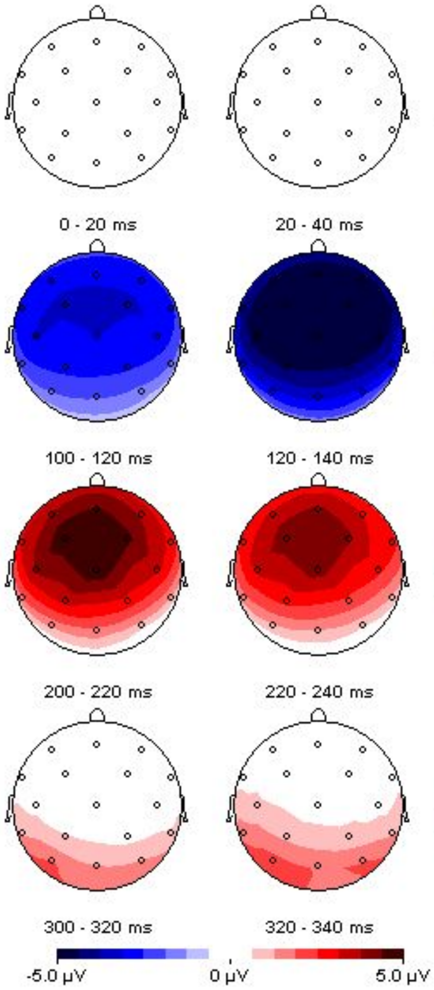

Fig.3.10 The scalp topographies of grand-averaged event-related potentials elicited by bimodal audiovisual (AV)stimuli

Multisensory integrations were assessed in the differences between the sum of unisensory auditory and visual ERPs(A+V) and the ERPs elicited by multisensory (AV) stimuli. The statistical significant differences (p values) of the AV - (A + V) were measured at 30 electrodes from 50 to 400 ms after stimuli onset.

In audiovisual integration experiment, the significant integrations were found at four periods: (1) From 100 to 160ms, in this latency window positive [AV - (A + V)] amplitudes were significant at 14 electrodes covering central and occipital scalp region. The greatest differences were found at Pz, Cp1 and Cp2 ($p<.01$). (2) From 160 to 200ms. The positive differences between AV and (A + V) were found at 16 electrodes covering central and occipital scalp region. The greatest differences were found at CP2, Pz, O1 and Oz ($p<.01$). (3) From 200 to 240ms, in this latency window the audiovisual positive interaction was the most significant, the greatest differences were found at Pz, O1 and Oz ($p <.001$). (4) Frontal-central areas from 280to320ms. The negative differences were found at F3, Fz, FC1, FC2 and Cz ($p<.05$).

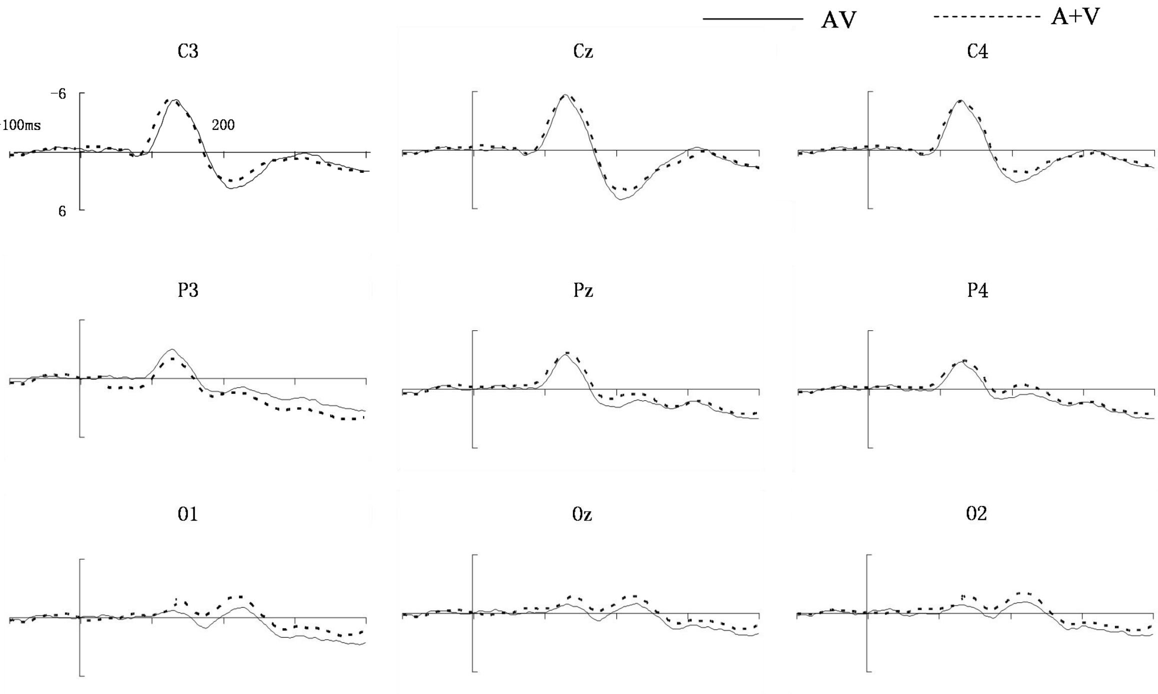

Fig. 3.11 Grand average event-related potentials (ERPs) to biomodal AV and the sum of unimodal A and V stimuli(A+V) at F3, Fz, F4, C3, Cz, C4, P3, Pz and P4.

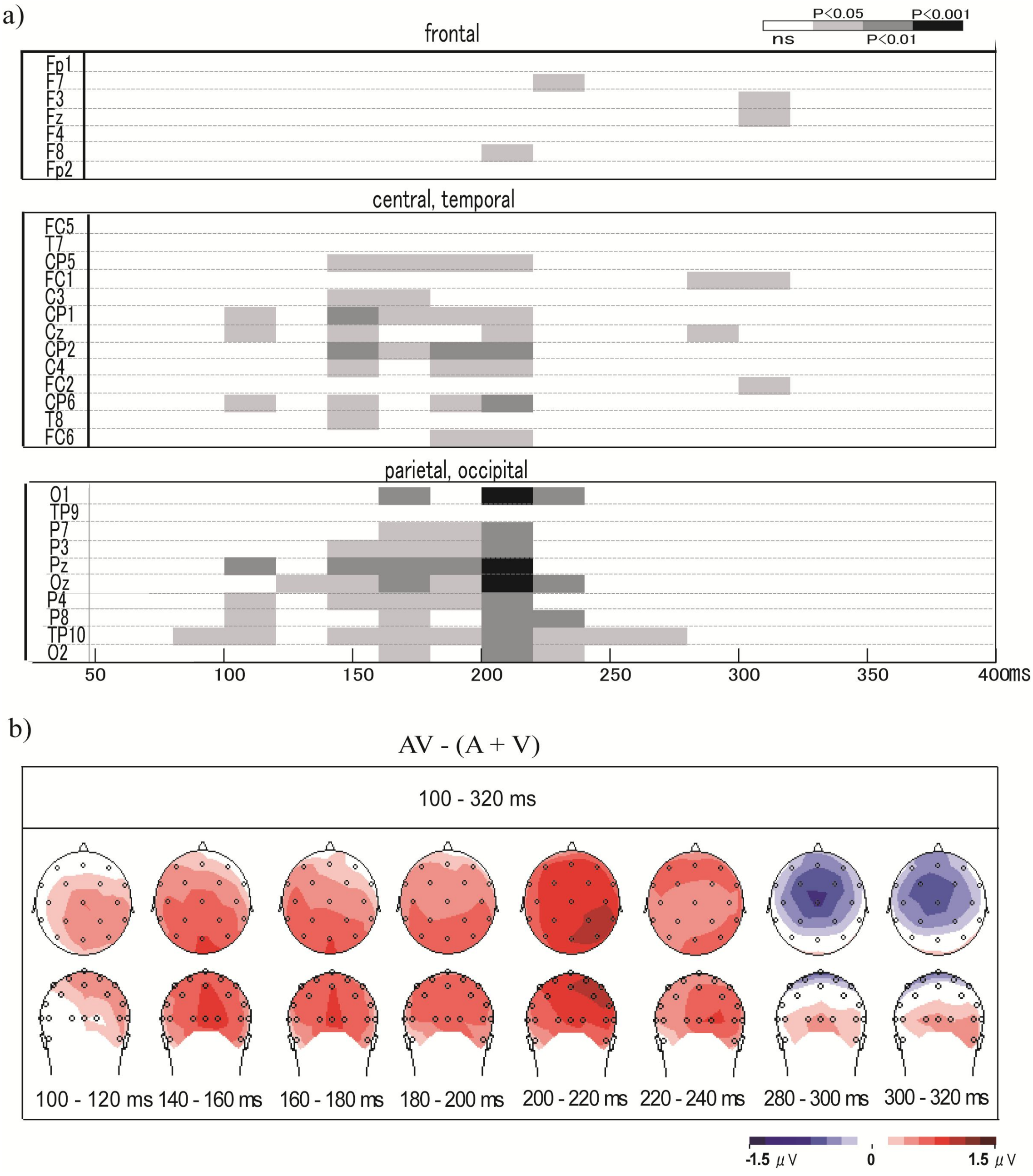

Fig.3.12 a) Spatio-temporal representation of the statistical significant differences (p values) of AV - (A + V) measured at 30 electrodes(repeated-measures analysis comparing the amplitude of AV and (A + V) at each latency).Time is plotted on the x axis from 50 to 400 ms. Electrodes are plotted on the y axis. b) Topographical maps of the voltage amplitudes for AV - (A + V) from100 to 120 ms, 140~240ms,280~320ms after stimulus onset.

3.4 Discussion

This study revealed behavioural and electrophysiological evidence for the influence of task-irrelevant auditory stimuli on audiovisual integration. Behavioural data showed that the RTs and HRs to audiovisual stimuli were significantly faster and more accurate than unimodal visual condition. This finding is consistent with some previous studies, and supported that task-irrelevant sound improved the responses to visual detection [1.1, 1.2] The ERP responses occurring during the first 200 ms following presentation of the stimulus were thought to be mediated by early sensory processing, whereas activity after 200 ms was thought to be mediated by late cognitive processing[1.9,3.3].

From 100 to 160ms

The early interaction was observed from 100 to 160ms, the latency and topographic distribution similar to those of unimodal auditory N1 wave, from 100 to120ms similar to the ascending limb of N1, from 140 to 160ms similar to the descending limb of N1. The opposite polarity of the topography suggests that the effects of interaction could decrease amplitude of the auditory N1 wave, indicating weaker activation of its generators in the auditory cortex for bimodal stimuli. [3.4].

From 160 to 240ms

The most prominent interaction was observed in this latency range at central and occipital electrodes sites. In the unimodal responses, this latency window corresponded to the ascending limb of the auditory P2 wave and to the visual N1 wave. The visual N1 wave is aroused at extra-striated, and associated with visual discrimination processes [3.5].

From 280 to 320ms

In this latency window, the negative interaction effects were found over frontal-central areas. Previous studies of audiovisual interactions have found analogous effects at frontal-central areas in which both selective attention task [3.6-3.9].

Our study provides evidence that multimodal integration can be generated even if the auditory stimulus was task- irrelevant. Four ERP components related to audiovisual integration were observed. Those finding confirms the main neural activity of audiovisual integration. In addition, our results suggested that the audiovisual integration elicited by task- irrelevant auditory could occur at both early and later processing.

3.5.Conclusion

Our study provides evidence that multimodal integration can be generated even if the auditory stimulus was task- irrelevant. Four ERP components related to audiovisual integration were observed. Those finding confirms the main neural activity of audiovisual integration. In addition, our results suggested that the audiovisual integration elicited by task- irrelevant auditory could occur at both early and later processing.

References

[3.1] Wallace, M.T. et al. Multisensory integration in the superior colliculus of the alert cat. J. Neurophys. 1998, 80, 1006–1010

[3.2] W. David Hairston, Donald A. Hodges, Jonathan H. Burdette and Mark T. Wallace. Auditory enhancement of visual temporal order judgment. Neuroreport 2006; 17: pp. 791–795.

[3.3]McDonald JJ, Ward LM. Involuntary listening aids seeing: evidence from human electrophysiology. Psychol Sci 2000; 11:167–171.

[3.4] Vidal J, Giard MH, Roux S, Barthelemy C, Bruneau N., Cross-modal processing of auditory-visual stimuli in a no-task paradigm: a topographic event-related potential study. Clin Neurophysiol., 2008, 119: 763-771.

[3.5] Vogel EK, Luck SJ. The visual n1 component as an index of a discrimination process. Psychophysiology 2000;37:190–203.

[3.6] Gemma A. Calvert, Peter C. Hansen , Susan D. Iversen, and Michael J. Brammer. Detection of Audio-Visual Integration Sites in Humans by Application of Electrophysiological Criteria to the BOLD Effect. NeuroImage 14, 2001;427–438 .

[3.7] Talsma D, Doty TJ, Woldorff MG. Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? Cereb Cortex 2007; 17:679–690.

[3.8] Jinglong Wu, Qi Li, Ou Bai, and Tetsuo Touge, Multisensory interactions elicited by audiovisual stimuli presented peripherally in a visual attention task: a behavioral and event-related potential study in humans. Journal of Clinical Neurophysiology.2009; 26, pp.407-413.

[3.9] Qi Li, Jinglong Wu and Tetsuo Touge, Audiovisual interaction enhances auditory detection in late stage: an event-related potential study. Neuroreport 2010; 21: pp. 173–178.

Chapter 4

The effects of temporal asynchrony on audiovisual integration

Summary

Many studies have shown that temporal asynchrony of visual-auditory stimuli can influence the interaction between visual and auditory stimulus, however, the multisensory mechanisms of asynchrony inputs were not well understood. In this study, the visual and auditory stimuli onset asynchrony (SOA= ±400 ms, ±150 ms, 0 ms), only the visual stimulus was attended. Behavioral data and Event-Related Potentials (ERPs) were recorded. The results found that when the SOA was -150ms, the reaction time was the fastest and hit rate was the highest, and the N1 latency of biomodal AV was earlier than the sum of unimodal A and unimodal V in auditory preceding condition.

4.1 Introduction

One of the basic functions of the human brain is to integrate a variety of sensory information. Previous studies have shown that reaction is more quickly and accurately to bimodal audiovisual stimuli compared with either unimodal visual or auditory stimuli [1.1, 1.2]. Spatially and temporally coincidence are two fundamental factors that facilitate the integration of audiovisual stimuli [4.1, 4.2]. Many previous studies have been investigated the perceptual processes underlying audiovisual integration when visual and auditory stimuli are presented simultaneously at same location [1.9,3.7,3.8].

However, in real-life, audiovisual stimuli onset asynchrony is ubiquitous phenomenon. For example, it’s well known that the phenomenon was seeing lightning before hearing the associated thunder. When the auditory and visual stimuli are input asynchronously, perceptual reports regarding the temporal characteristics of visual stimuli tend to be biased toward the temporal characteristics of the auditory stream [1.13,4.3]. For example, temporal disparity was easier in arousing the illusion if the visual stimulus was presented before the auditory stimulus by about 50-100ms [4.3]. There is a relatively time window of integration as large as 250 ms, in which stimuli from different modalities tend to be integrated into a single unified percept [4.4,4.5]. However, the effects of visual detection by temporal alignment are not well understood.

Present study designed a visual detection task. The stimuli onset asynchrony (SOA) between auditory and visual stimuli (SOA = ±400 ms, ±150 ms, 0 ms), unimodal visual (V), unimodal auditory (A) and AV (5 types) stimuli were presented randomly. The participants only attend to the visual modality. This experiment used behavioral and Event-Related Potential (ERP) measured to investigate the effects of 5 types SOA on visual detection task.

4.2 Methods

4.2.1 Subjects

Thirteen healthy students (23 to 31 years, mean: 25.9 years, 5 females) from Okayama University participated in behavioral and ERP experiment after receiving a full explanation of the purpose and risks of the research. All subjects had normal or corrected-to-normal vision and normal hearing and right handed.

4.2.2 Stimuli

Unimodal visual stimuli consisted of 0° (V standard stimulus) and 90° (V target stimulus) Gabor gratings (2°diameter, 2 cycles/degree) (Fig.1), which could appear 4° below the fixation point during 23ms. The contrast of these ranged from 3 to 7% in units of Michelson contrast [(max−min)/(max+min)], with max and min being the maximal and minimal values of the Gabor patch. We conducted a pre-experiment prior to all experiments, in order to determine one intermediate contrast level so that the accuracy in detecting the stimuli was over 80%.

Unimodal auditory (A) stimuli consisted of 1000 Hz tone, presented at a sound pressure level of 70 dB for duration of 40 ms (including linear rise and fall times of 5 ms). The auditory stimuli were presented through two speakers that were placed on both sides of the display.

Bimodal audiovisual (AV) stimuli consisted of visual and auditory stimuli onset asynchrony (SOA). The SOA are −400, −150, 0, 150 and 400ms. The "−" means auditory stimulus was presented earlier than visual stimulus. "SOA=0" means visual and auditory stimuli onset synchronously.

4.2.3 Task and procedure

Experiments were conducted in a dark and sound-attenuated room. Visual stimulus was presented on a 17-inch display. The background luminance was set to 2.7cd/m^2 and was perceived as neutral grey. Participants viewed the display from a distance of 57 cm, with their head placed on a custom-made chin rest. Each trial started with the presentation of a fixation plus at the centre of the screen, followed by a random interval (800 to 1400 ms), then the A, V or AV(one of five types) stimuli was presented with equal probability(Fig.4.1).

The participants' task was to only attend to the visual modality and only respond to the visual target stimuli. In stimuli of each type, target stimuli were 20%.

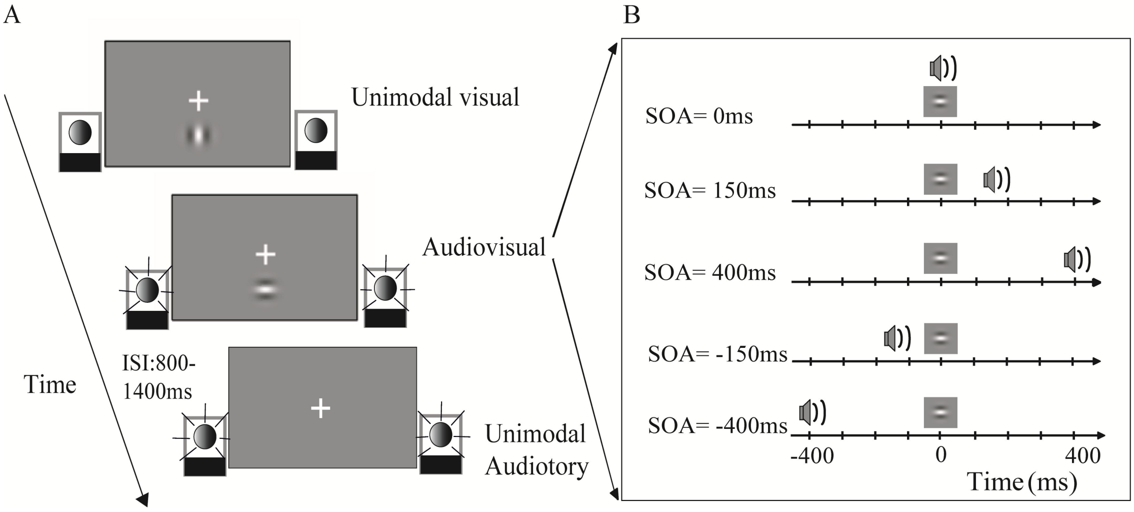

Fig. 4.1 A: Example stimuli in the experiment. The unimodal visual stimulus was presented on a display at an angle of 4°below the fixation point. And then audiovisual was shown. Then unimodal auditory stimulus was presented through two speakers. The interstimulus interval (ISI) is 800 to 1400ms. B: 5 types audiovisual stimuli. SOA = 0ms: visual and auditory stimuli onset synchronously; SOA = 150ms: auditory stimulus delayed by 150ms relative to the onset of the visual; SOA = 400ms: auditory stimulus delayed by 400ms relative to the visual; SOA = -150ms: visual stimulus delayed by 150ms relative to the auditory; SOA = - 400ms: visual stimulus delayed by 400ms relative to the auditory.

4.2.4 Data acquisition and analysis

Stimulus presentation was controlled by a personal computer running Presentation software (Version 0.61). Reaction time (RTs) and hit rates (HRs) were subjected to t-test for the different types of target stimuli

EEGs were recorded from 32 electrodes using BrainAmp MR plus (Gilching, Germany), mounted on an electrode cap (Easy cap, Herrsching Breitbrunn, Germany) according to the 10/20 international standard system. All signals were referenced to the linked two ear electrodes. All electrode impedance was kept below 5kΩ. Horizontal eye movements were recorded from the outer canthus of the left eye. Vertical eye movements and eye blinks were recorded from an electrode below the left eye. Raw signals were digitized with a sample frequency of 500Hz with a 60Hz notch filter and all data were stored digitally for off-line analysis. An artifact criterion over ± 80μV was used at all channels to reject trials with noise transients. The data were then averaged for each stimulus type following digital filtering using a band-pass filter of 0.01–30Hz. After filtering, final grand-averaged data were obtained across all subjects for each stimulus type.

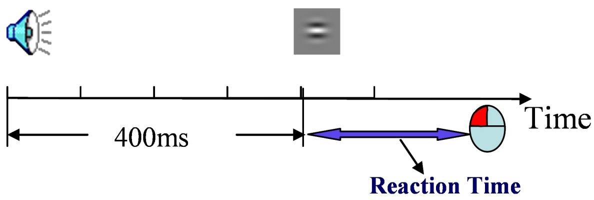

Fig.4.2 Recording method of reaction times

4.3 Result

4.3.1 Behavioral Results

Reaction times (RTs) were recorded from V target stimulus onset to pressing the response key, and means of group were showed in Fig. 4.3. RTs to AV target stimuli of SOA = -150ms [t (12) = 4.3, $P < 0.01$] and SOA = 0 [t (12) = 3.7, $P < 0.01$] were significantly faster than that to unimodal visual (V) condition. RTs of SOA=400ms [t (12) = -3.4,$P < 0.01$] were significantly slower than for unimodal V condition. RTs of unimodal V [t (12) = 3.7, $P <0.01$], SOA=150ms [t (12) = -4.0, $P <0.01$] and SOA = 400ms [t (12) = -4.4, $P < 0.01$] were significantly slower than for simultaneous audiovisual condition (SOA = 0).

The hit rates (HRs) of AV target stimuli in SOA =-150ms [t (12) = -3.7, $P < 0.01$], SOA = 0 [t

(12) = -4.5, P < 0.01], SOA = 150ms [t (12) = -5.0, P < 0.001], SOA = 400ms [t (12) = -3.5, P < 0.01] conditions were more accurate than these to unimodal V condition. There was no statistical difference on accuracy between unimodal V and AV of SOA=-400ms condition [t (12) = -1.0, P = 0.33].

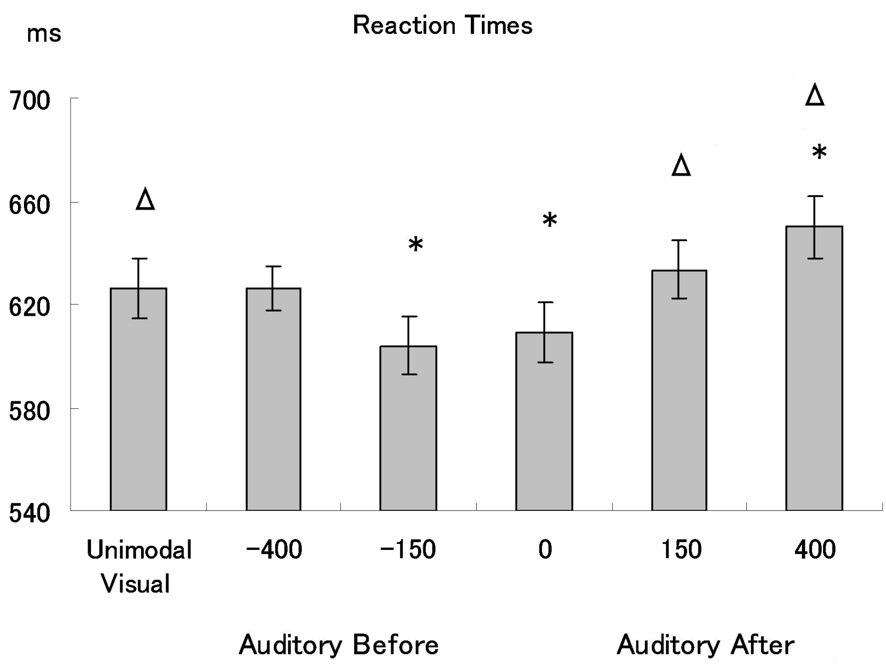

Fig. 4.3 Average response times (RTs). RTs are in millisecond. Unimodal visual denote conditions in which only visual sitimulus was presented. Negative SOA (i.e.-400ms) correspond to the auditory stimulus presented before the visual stimulus. Asterisks (*) denote conditions in which RTs are significantly (P<0.01) different to unimodal visual. Triangle signs (Δ) denote conditions in which RTs are significantly different (P<0.01) from the simultaneous (i.e. SOA= 0ms) audiovisual condition.

The HRs to AV target stimuli of SOA = 400ms were more accurate than these to unimodal V condition, but the RTs slower than unimodal V condition. In order to ensure that the effect on RTs did not related to participant's changes in accuracy, the correlation analysis was measured between Hits and RTs (r = 0.088, P = 0.774). The result that no significant correlation were found.

So, it suggested that the auditory stimuli after 400ms visual stimuli onset enhanced the visual detection(Fig.4.4).

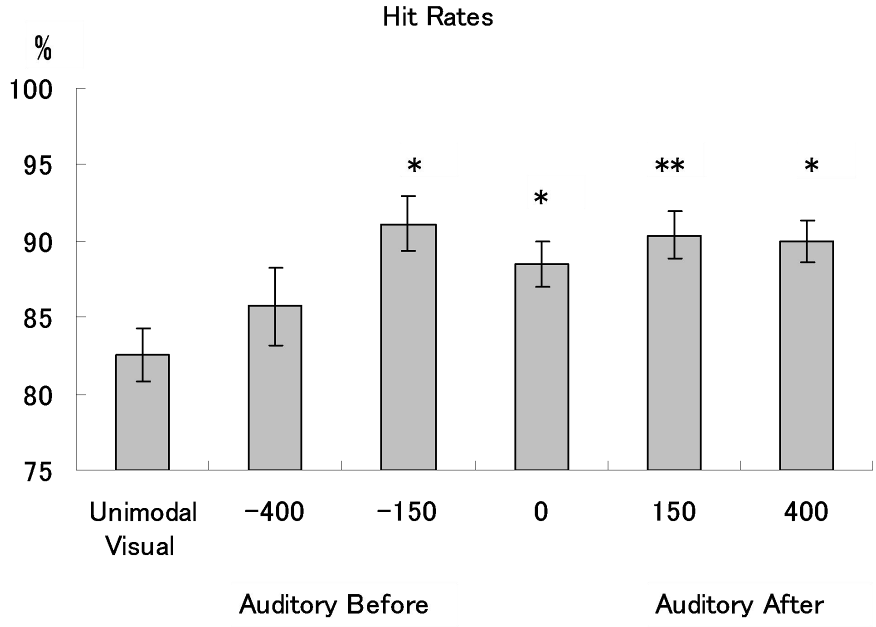

Fig.4.4 Hit rates (HRs) in the detection of each stimulus type. Asterisks (*) denote conditions in which HRs are significantly more accuracy than for the unimodal visual. (* $P<0.01$, ** $p < 0.001$)

The most significant enhancement was SOA = -150ms condition, which RT was the fastest and HR was the most accurate. Then we compared ERPs of AV in SOA = -150ms condition and the sum of unimodal V and unimodal A.

4.3.2 ERPs Results

ERP s results of SOA= 0ms

Only the ERPs elicited by standard stimuli were analyzed. Continuous EEG signals were divided into epochs from 100ms before stimulus onset to 400ms after stimulus onset. Audiovisual integration were detected by comparing the ERPs to AV stimuli with the sum of the ERPs to A and V stimuli (AV vs. A + V).

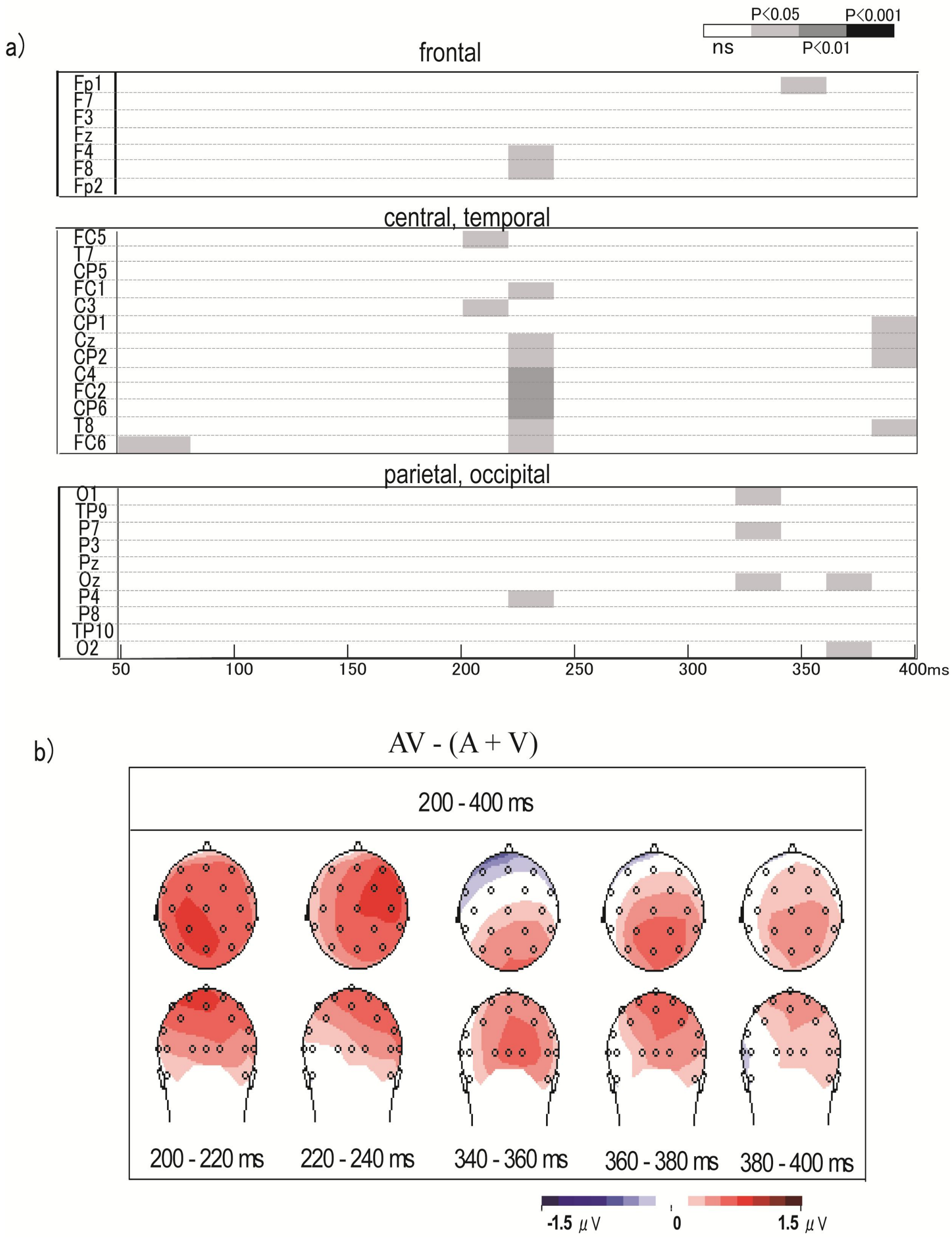

Fig.4.5 Unreliable stimuli experiment., a) Spatio-temporal representation of the statistical significance (p values) of the AV - (A + V) diff erences measured at 30 electrodes(repeated-measures analysis comparing the amplitude of AV and (A + V) at each latency).Time is plotted on the x axis from 50 to 400 ms. Electrodes are plotted on the y axis. b) Topographical maps of the voltage amplitudes for AV - (A + V) from200 to 240 ms,340~400ms after stimulus onset.

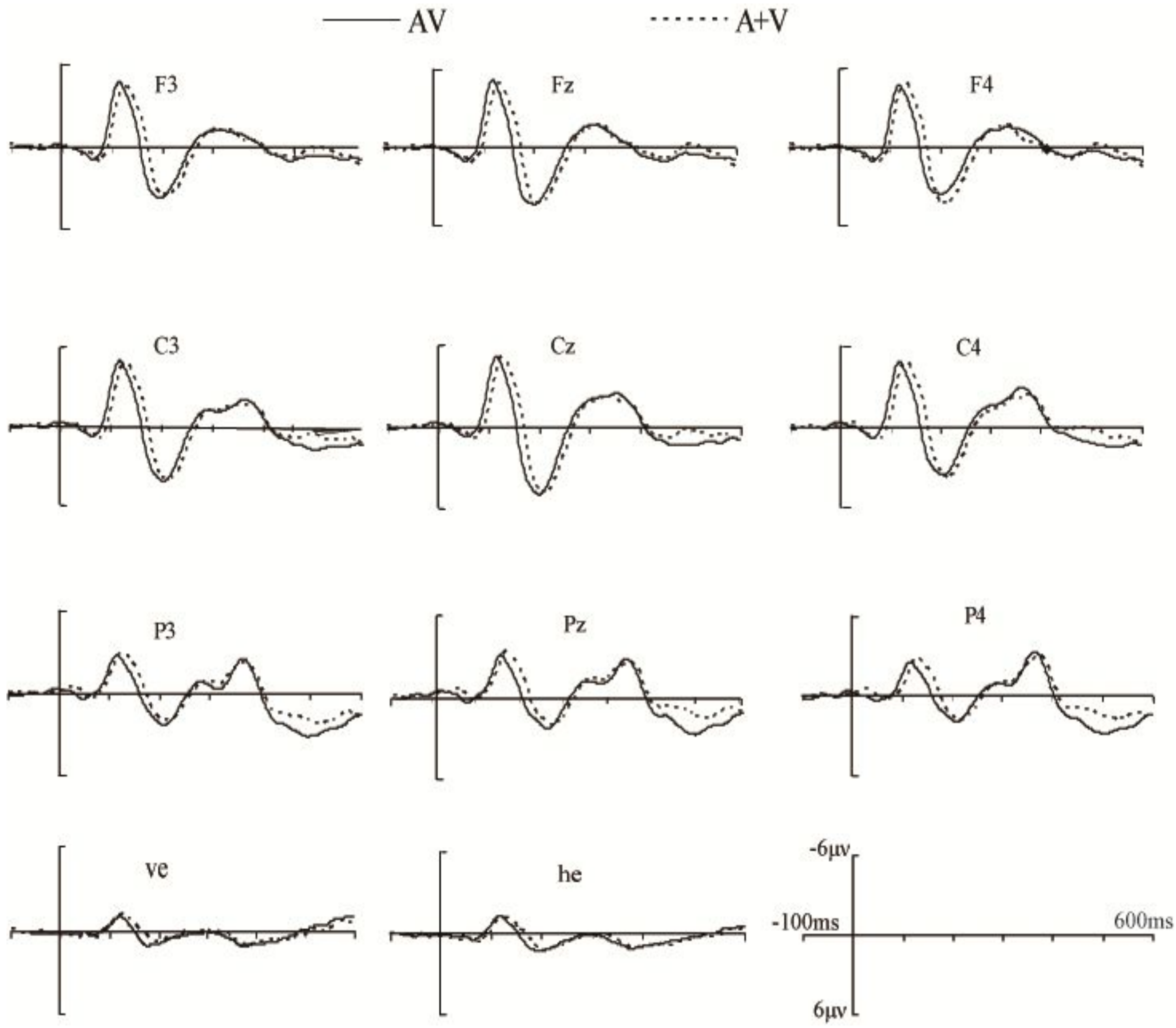

Fig.4.6 Grand average event-related potentials (ERPs) to biomodal AV (SOA = -150ms) and the sum of unimodal A and V stimuli (A+V) at F3, Fz, F4, C3, Cz, C4, P3, Pz and P4

The significant differences between AV and (A + V) were found at three areas: (1) right-temporal areas around 220 -240 ms: in this latency window [AV-(A+V)] amplitudes were significant at T8(p=0.004)(Fig.4.4A); (2) occipital areas at around 340-380ms: the significant differences were observed at Pz(p=0.001), P4(p=0.027),Oz(p=0.025), O2(p=0.044)(Fig.4.4B); (3)central areas around 340-380ms: the significant differences were observed at Cz (p=0.031), CP1 (p=0.012), CP2(p=0.005) (Fig.6C).

The reason of enhancement visual detection by AV (SOA= -150ms) stimuli was assessed in the differences between the sum of unimodal auditory and visual ERPs(A+V) and the ERPs elicited by biomodal (AV) stimuli of SOA= -150ms. The ERPs which were elicited by AV and A stimuli were recorded from -100 to 600ms.

The ERPs of unimodal V stimuli were recorded from -250ms (before V stimuli onset) to 450ms (after V stimuli onset). The ERPs of AV and (A+V) presented a negative N1wave (peaking at around 100ms), as showed in Fig.4.4. The N1 latency of AV was significant earlier than (A+V). The statistical result of AV and (A+V) was showed in table 1, the differences were observed at 24 electrodes excluding O1, O2, Oz and T7.

4.4 Discussion

In this study, we analyzed the behavioral results and ERPs recordings which were elicited by temporal asynchronous visual-auditory stimuli. The behavioral results showed that the responses to temporal asynchronous AV stimuli were more accurate than unimodal V. In other words, the temporal asynchronous AV stimuli enhanced the visual detection [1.13,4.6].

In SOA= 0ms condition, behavioral data showed that RTs to simultaneous audiovisual stimuli were significantly faster than visual only, even though the subject could not found the cognitive rules connecting the two stimuli. This finding is consistent with some previous studies, and supported that task-irrelevant sound improved the responses to visual detection [4.1,4.2].

Table4.1 Statistical Results of N1 latency (A+V vs AV) for the SOA = -150ms condition from 28 electrodes. All T-tests were conducted with (1, 12) degrees of freedom.

Electrodes	A+V		AV		F	p
	Mean	Std. Error	Mean	Std. Error		
C3	124.92	1.46	114.46	1.83	100.89	0.000
C4	126.77	1.63	114.77	1.81	96.83	0.000
Cz	127.54	1.65	115.23	1.90	229.94	0.000
CP1	125.69	1.58	114.15	1.78	89.05	0.000
CP2	126.77	1.73	114.62	1.86	91.56	0.000
CP5	123.54	2.22	115.54	3.03	5.16	0.042
CP6	130.00	2.77	114.92	2.28	38.34	0.000
Fz	127.38	1.58	114.92	1.53	176.53	0.000
F3	126.46	1.92	115.69	1.88	69.50	0.000
F4	128.62	2.19	116.62	1.90	108.00	0.000
F7	123.54	1.80	114.46	2.80	17.93	0.001
F8	125.54	2.52	116.00	2.36	29.99	0.000
FC1	126.15	1.61	115.69	1.55	78.59	0.000
FC2	127.38	1.58	115.08	1.69	198.96	0.000
FC5	126.00	2.28	113.85	2.19	23.14	0.000
FC6	126.92	2.19	115.23	1.94	45.30	0.000
FP1	127.54	2.52	117.38	1.82	31.19	0.000
FP2	127.23	2.82	116.62	2.14	12.97	0.004
P3	126.15	2.27	117.08	3.32	6.97	0.022
P4	128.92	3.56	114.46	2.31	26.12	0.000
P7	125.08	4.01	111.38	4.38	6.06	0.030
P8	135.23	3.35	117.69	5.60	12.87	0.004
Pz	127.08	2.35	115.85	2.59	17.38	0.001
T7	128.46	3.60	119.54	4.71	4.21	0.063
T8	131.69	3.47	118.62	3.70	8.39	0.013
O1	127.08	4.64	117.23	4.90	3.01	0.108
O2	127.08	4.71	117.23	4.93	2.08	0.175
Oz	126.15	4.83	117.69	4.85	2.70	0.126

Responses occurring during the first 200 ms following presentation of the stimulus (POS) were thought to be mediated by early sensory processing, whereas activity after 200 ms was thought to be mediated by late cognitive processing [3.7,3.8].Our ERPs results did not show any effects of multimodal AV interaction in the time period up to 200 milliseconds after POS, as found in previous divided-attention task [3.7]. As shown in Figu6, all the different values of [AV-(A+V)] were positive. The positive values in the [AV-(A+V)] differences at these sites were therefore due to smaller amplitudes of the positive potentials in (A + V) than in AV conditions.

A) The difference over the right temporal area at around 220-240ms

The first effect of AV interaction was observed over a right temporal area in scalp topography around 220-240ms. The spatial location of ERPs in this study is similar to Giard and Peronnet's [1.19].The integration effect was most likely elicited by the anteriorpart of the right temporal lobe and/or the inferior lateral prefrontal cortex.

B) The difference over the occipital and central area at around 340-380ms

areas around 340 to 380 ms. Few assessed the brain activity after 200ms following presentation of the stimulus. The integration effect results suggested that the activities of visual cortex were increased for bimodal stimuli after [4.3].

In SOA = -150ms condition we observed an audiovisual integration effects at occipital and central Responses were the fastest and the most accurate when the A stimuli preceded V stimuli about 150ms (SOA = -150ms). The auditory stimuli which were presented earlier than visual stimuli were regarded as cue. The temporal relationship of auditory cues to visual targets can markedly influence the accuracy and the speed of visual detection [4.6]. From the ERPs waveforms, we found the N1 latency of AV was significant earlier than (A+V) at frontal, central and parietal areas. The preceding auditory cue enhanced attention to the following visual stimuli, even though the auditory stimuli were task-irrelevant. The attention could enhance or modulate multisensory integration [1.9].

In SOA = -400ms condition we did not found significant statistical difference. The reason is that audiovisual integrations depended on a time window as large as 250 ms delays between auditory and visual stimuli [1.20,4.4]. In SOA = 0 condition the RT was faster and HR was more accurate than Unimmodal V stimuli. It was proved that simultaneous audiovisual stimuli aroused significant integration and enhanced the cognition to visual stimuli [1.1,4.2,3.7,3.8]. In SOA = 150, 400ms condition the auditory stimuli improved the HR. That explained that the posterior auditory provided the temporal profile and enhanced the attention to visual stimuli [4.6], although the auditory stimuli were task irrelevant.

4.5.Conclusion

This study revealed behavioral and ERPs evidence for the effects of audiovisual temporal alignment on visual detection, and found that the temporal alignment enhanced the visual detection and could enhance or modulate multisensory integration.This enhancement was the most prominent when the A stimuli preceded V stimuli about 150ms (SOA = -150ms).

References

[4.1] Lewald J, Guski R.. Cross-modal perceptual integration of spatially and temporany disparate auditory and visual stimuli. *Cogn Brain Res.* 2003;16: pp. 468--478.

[4.2] Michael Lippert, NK. Logothetis, C.Kayser, Improvement of visual contrast detection by a simultaneous sound. *Brain Res.* 2007; 1173: pp. 102–109.

[4.3] Slutsky DA, Recanzone GH. Temporal and spatial dependency of the ventriloquism effect. *Neuroreport* 2001; 12: pp. 7–10.

[4.4] Charles Spence and Sarah Squire. Multisensory Integration: Maintaining Dispatch the Perception of Synchrony. *Current Biology.* 2003; Vol. 13, pp. 519–521.

[4.5]Jorg Lewald, Walter H. Ehrenstein, Rainer Guski. Spatio-temporal constraints for auditory–visual integration. *Behavioural Brain Research* 2001; 121, pp. 69–79.

[4.6] W. David Hairston, Donald A. Hodges, Jonathan H. Burdette and Mark T. Wallace. Auditory enhancement of visual temporal order judgment. *Neuroreport* 2006; 17: pp. 791–795.

Chapter 5

Temporal Reliability of Auditory Stimuli Modulates Audiovisual Integration

Summary

This chapter compared the integration elicited by synchronous audiovisual stimuli between the reliable and unreliable conditions. In the reliable auditory stimuli condition, auditory and visual stimuli were always presented simultaneously; in the unreliable condition, stimuli were presented with a stimulus onset asynchrony of 0 ms, 150 ms, or 400 ms. Audiovisual integration in the reliable condition occurred around 140 – 200 ms, covering the parietal and occipital regions; in the unreliable condition, integration occurred later, around 320 – 340 ms, and covered the occipital region. These results show that the temporal reliability of auditory stimuli can modulate audiovisual integration at the early sensory processing stage.

5.1 Introduction

In everyday life, the human brain is capable of combining information from multiple sensory modalities. Information from different sensory organs is often efficiently merged to form a unified and robust percept. Previous studies have shown that audiovisual stimuli are detected and discriminated more accurately than either visual or auditory stimuli alone. This process is referred to as "audiovisual integration" [1.1,5.1].

Whether the auditory and visual stimuli are combined or not, it is relies on the stimulative accordance of physical and cognitive factors [5.2]. The physical factors contain temporal, spatial, intensity of stimulus [5.2,5.3]. The cognitive factors contain whether the subject thinks that bimodal stimuli should occur together [5.2], whether bimodal stimuli accord semantically [1.10].

The effect of physical factors on audiovisual interaction has been well studied [5.3-5.5]. Spatially and temporally coincident audiovisual stimuli are detected more easily and more quickly than unimodal stimuli [5.4]. In real-life, audiovisual stimuli onset asynchrony is ubiquitous phenomenon. For example, it ’ s well known that the phenomenon was seeing lightning before hearing the associated thunder. When the auditory and visual stimuli are input asynchronously, perceptual reports regarding the temporal characteristics of visual stimuli tend to be biased toward the temporal characteristics of the auditory stream [5.5-5.7]. For example, temporal disparity was easier in arousing the illusion if the visual stimulus was presented before the auditory stimulus by about 50-100ms [5.5]. There is a relatively time window of integration as large as 250 ms, in which stimuli from different modalities tend to be integrated into a single unified percept.

However, the cognitive factors seem to have a stronger impact on the sensory interaction than the physical association between auditory and visual stimuli [5.2]. In many previous studies audiovisual stimuli are presented always simultaneously [5.6,1.12]. The subjects can find that the auditory stimulus provided reliable and fixed temporal relative to visual stimuli, therefore the auditory stimuli are regarded as informative sound. An informative sound can reduce the uncertainty about the timing of the visual display, and significantly improve detection rates. If the subject cannot find the fixed time information of auditory stimuli ,even though audiovisual stimuli are presented simultaneously, the auditory stimuli are considered as a completely redundant uninformative sound. A completely redundant sound cannot alter detection rates, but it can speed reaction times [5.2].

Temporal coincidence of auditory and visual stimuli is an important factor for audiovisual integration[5.2,5.3]. Studies have revealed that audiovisual integration is most significant when the auditory and visual stimuli are presented simultaneously [5.3,1.10]. In those studies, visual and auditory stimuli had onset synchrony; auditory stimuli could thus provide reliable temporal information for visual perception. When the auditory and visual stimuli were presented with a stimulus onset asynchrony (SOA) [5.4,1.12], participants failed to find the temporal relationship between the visual and auditory stimuli. In this case, the temporal information provided by the auditory stimuli was thus unreliable. Some researchers have investigated the effects of signal reliability on multimodal integration [5.5,5.6]. Jean-Pierre Bresciani and Marc O. Ernst found that lowering the reliability of an auditory stimulus can decrease the effect of audition on touch. Qiangliu et al. showed that audiovisual integration occurs more effectively with high-reliability visual stimuli than with low-reliability stimuli [5.7]. However, it is not clear whether the temporal reliability of auditory modalities can modulate audiovisual integration.

The aim of present study was to compare the differences between audiovisual integration elicited by reliable versus unreliable auditory stimuli through behavioral and event-related potentials (ERPs) experiment. In the reliable auditory stimuli condition, the visual and auditory stimuli were always presented simultaneously. In the unreliable auditory stimuli condition, the auditory stimulus was randomly presented before, in synchrony with, or after the visual stimulus; the temporal information provided by the auditory stimulus was thus unreliable. The participants were instructed to focus on the visual modality and to ignore the auditory modality. We tested how the temporal reliability of the auditory stimulus affected audiovisual integration.

5.2 Methods

5.2.1 Subjects

Thirteen healthy volunteers (23 to 31 years, mean: 25.9 years, 5 females) from the Okayama University participated in the experiment. All participants were right-handed and had normal or corrected-to-normal vision and normal hearing. After receiving a full explanation of the purpose and risks of the research as per the protocol approved by the institutional research review board, subjects gave written informed consent for all studies.

5.2.2 Stimuli

Experiments were conducted in a dark and sound-attenuated room. Visual stimuli were presented on a 17-inch display. The background luminance was set to 2.7cd/m^2 and was perceived as neutral gray. Participants viewed the display from a distance of 57 cm, with their head placed on a custom-made chin rest. Unimodal visual (V) stimuli consisted of vertical (standard stimulus) and horizontal (target stimulus) Gabor gratings (2°diameter, 2 cycles/degree), which appeared 4° below the fixation point for a duration of 23ms. The contrast of these ranged from 3% to 7% in units of Michelson contrast: (max−min)/(max+min), with the maximal (max) and minimal (min) values of the Gabor patch. We conducted a preparation experiment prior to all experiments in order to find an intermediate contrast level in which the accuracy in detecting the stimuli was 80%.

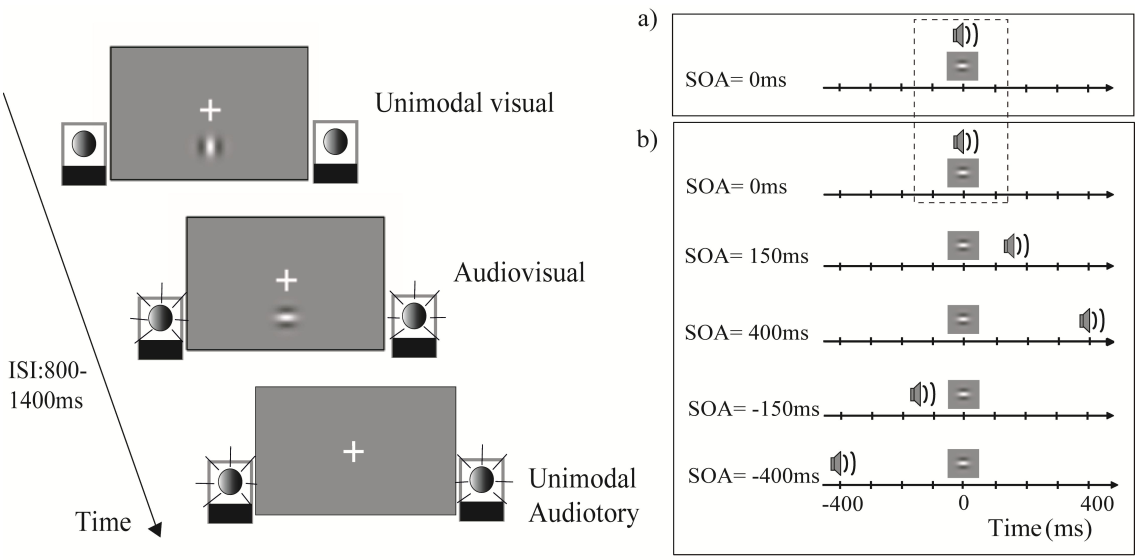

Fig. 5.1 General layout of the paradigm and stimuli. The visual stimulus was presented on a display at an angle of approximately 3°below the fixation point. The auditory stimuli were presented through two speakers placed on the left and right of the display. (a) Auditory and visual stimuli were presented simultaneously (SOA = 0) in the reliable auditory stimuli condition. (b) Audiovisual stimulus onset asynchrony (SOA = ±400 ms, ±150 ms, or 0 ms) in the unreliable auditory stimuli condition. We compared the integration elicited by SOA = 0 ms between the reliable and unreliable auditory stimuli conditions. The participants' task was to attend only to the visual modality and to respond only to the visual and audiovisual target stimuli.

Unimodal auditory (A) stimuli consisted of 1000 Hz sinusoidal tones, presented at a sound pressure level of 70 dB through two speakers that were placed on either side of the display, for a duration of 40 ms, including linear rise and fall times of 5 ms.

In the reliable auditory stimuli condition, the bimodal audiovisual (AV) stimulus was always composed of the simultaneous presentation of unimodal A and V stimuli. In the unreliable auditory stimuli condition the unimodal A and V stimuli were presented with a stimulus onset asynchrony (SOA = ±400 ms, ±150 ms, or 0 ms, as shown in Figure 1b).

5.2.3 Task and procedure

Each block had a 3000 ms fixation period followed by the test stimulus. V, A and AV stimuli were presented randomly. The interstimulus interval varied randomly between 800 and 1400 ms. The reliable auditory stimuli condition contained three blocks, the standard stimuli for each type (V, A, and AV) were presented 183 times, and the target stimuli (V and AV) were presented 48 times. The unreliable auditory stimuli condition contained eight blocks, the standard stimuli for each type (V, A, and AV of 5 SOA types) were presented 184 times, and the target stimuli (V and AV of 5 SOA types) were presented 48 times. The participants' task was to attend only to the visual modality and to respond only to the V and AV target stimuli.

5.3 Data acquisition and analysis

5.3.1 Behavioral analysis

Stimulus presentation was controlled by a personal computer running Presentation software (Neurobehavioral Systems, Albany, CA). Reaction times (RTs) and hit rates (HRs) responding to the target stimuli were composed separately for each stimulus type. The RTs and HRs for the V and AV target stimuli were analyzed using a repeated-measures analysis of variance (ANOVA) with variances of stimulus modality (V, AV) and auditory reliability (reliable or unreliable condition). In the unreliable condition, only the simultaneous AV (SOA = 0) and unimodal visual (V) conditions were of interest, since these traces were identical to the reliable condition. Another SOA goal to avoid having the subject to find the cognitive rules connecting the auditory and visual stimuli.

5.3.2 Event-related potentials (ERPs) analysis

Electrophysiological signals were recorded using BrainAmp MR Plus (Gilching,Germany) via 32 electrodes mounted on an electrode cap (Easy Cap, Herrsching Breitbrunn, Germany) according to the 10/20 international standard system. All signals were recorded using the linked earlobes as a reference. All electrode impedances were maintained below 5kΩ. Horizontal eye movements were recorded from the outer canthus of the left eye; eye blinks and vertical eye movements were recorded from an electrode attached below the left eye. Raw signals were digitized at a sample frequency of 500 Hz with a 60 Hz notch filter and all data were stored digitally for off-line analysis. An artifact criterion of over 80μV was used at all channels to reject trials with noise transients. The data were then averaged for each stimulus type following digital filtering using a band-pass filter of .01 – 30 Hz. After filtering, final grand-averaged data were obtained across all subjects for each stimulus type. Only the ERPs elicited by standard stimuli were analyzed. Continuous EEG signals were divided into epochs from 100 ms before stimulus onset to 400 ms after presentation of stimuli (POS). Multisensory integration was assessed as a comparison between the sum of unimodal auditory and visual ERPs (A+V ERPs) and the ERPs elicited by biomodal stimuli (AV ERPs). Waveforms of AV ERPs were subtracted from those of A+V ERPs. Mean amplitudes were calculated for all electrodes at consecutive windows of 20 ms each between stimulus onset and 400 ms after presentation of the stimulus. The mean amplitude data were analyzed using repeated-measures analysis of variance with the within-subjects factors of modality (AV, A+V), time-window, and electrodes.

5.4 Results

5.4.1 Behavioral data

Reaction times (RTs) and hit rates (HRs) were computed separately for each stimulus type (Table 5.1). There was no significant difference in RTs associated with either of the two conditions (reliable and unreliable) [$F(1,12) = 0.057$, $p = .815$]. Analysis of the RTs showed a main effect of modality (AV and V) [$F(1, 12) = 15.423$, $p = 0.002$]. However, no significant interaction between modality and reliability were found the RTs [$F(1, 12) = 0.039$, $p = .846$].

Table 5.1 Response times are in milliseconds (msec). Hit rates are in percentage of correctly responded targets. Standard deviation values are given in parentheses.

Mean Response Times and Hit Rates

	Reliable condition	Unreliable condition
Response Times (ms)		
Visual (V)	623(57)	626(42)
Audiovisual (AV)	607(48)	609(42)
Hit Rates (%)		
Visual (V)	88.9(6.0)	82.6(6.0)
Audiovisual (AV)	91.8(3.8)	88.5(5.2)

HRs of the reliable auditory stimuli condition were more accurate than those of the unreliable auditory stimuli condition [$F(1,12) = 8.064$, $p = .015$], where a main effect of modality was significant in the HRs [$F(1,12) = 23.845$, $p < .001$]. No significant interaction between modality and reliability were found in HRs [$F(1, 12) = 2.624$, $p = .131$].

5.4.2 ERP waveforms

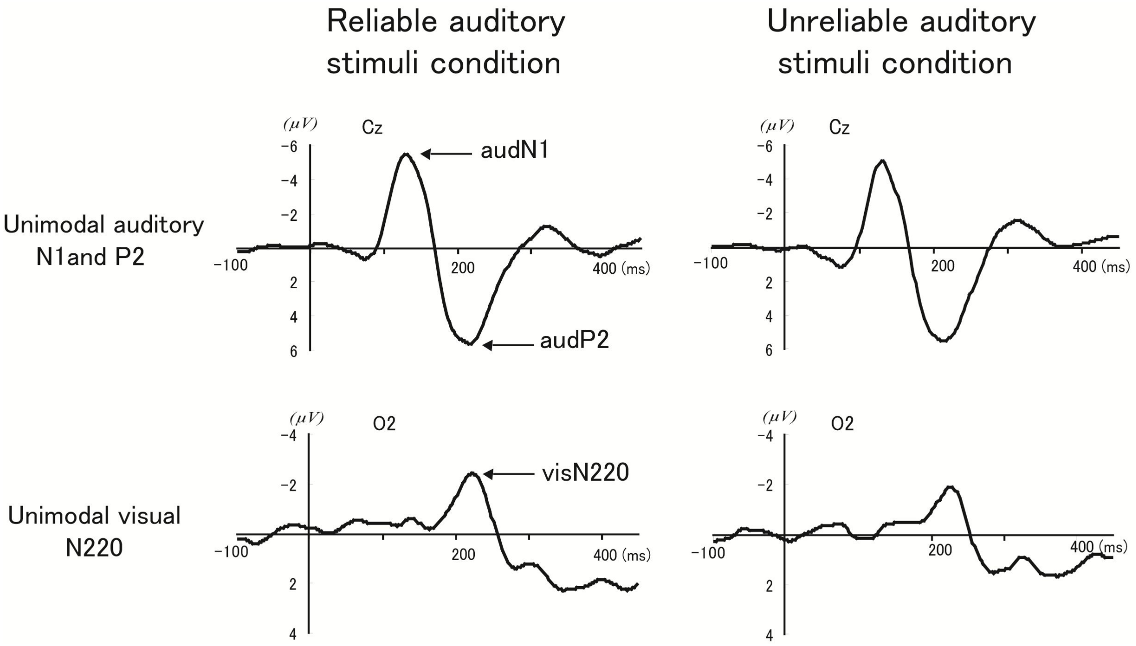

Fig. 5. 2 Grand-averaged event-related potentials (ERPs) at Cz and O2 for unimodal auditory stimuli and unimodal visual stimuli. The auditory ERPs included typical N1 (audN1) and P2 (audP2) waves, the maximum amplitude of which was at Cz. The visual ERPs presented a negative N220(visN220) wave peaking around 220 ms, the maximum amplitude of which was at O2.

The group-averaged ERP components elicited by unimodal auditory and unimodal visual stimuli are shown in Fig.5.2. The auditory ERPs included typical N1 and P2 waves. The negative N1 wave covering fronto-central areas peaked around 126 ms (-5.443 μV at Cz) in the reliable stimulus condition and 128 ms (-5.037 μV at Cz) in the unreliable stimulus condition. The positive P2 wave peaked around 214 ms (5.573 μV at Cz) in the reliable stimulus condition and 212 ms (5.535 μV at Cz) in the unreliable stimulus condition. The visual ERPs produced a negative N220 wave at occipital sites peaking around 222 ms (-2.406 μV at O2) in the reliable stimulus condition and 226 ms (-2.104 μV at O2) in the unreliable stimulus condition.

5.4.3 Reliable auditory stimuli's effect on Integration

In the reliable auditory stimuli condition, significant interactions between AV and (A+V) were found at 140-200 ms after POS. The amplitudes of the subtracted waveforms covering parietal and occipital regions around P3, Pz, P4, O1, Oz and O2 were significant (Fig.5.3A). The analysis yielded a significant effect of the modality [$F(1,12) = 20.48$, $p < .001$]. The difference in amplitudes between AV and A+V were apparent between 140 and 200 ms at P3 [$F(1,12)= 23.67$, $p< .001$](mean amplitude, AV-(A+V): 0.66 μV), Pz [$F(1,12) = 40.92$, $p < .001$] (mean amplitude: 1.23 μV) and P4 [$F(1,12) = 17.63$, $p = 0.001$] (mean amplitude: 0.56 μV). The occipital differences between AV and A+V were confirmed at O1 [$F(1,12) = 11.20$, $p = .006$] (mean amplitude: 0.95 μV), Oz [$F(1,12) = 6.06$, $p = .029$] (mean amplitude: 0.55 μV) and O2 [$F(1,12) = 6.52$, $p = 0.029$] (mean amplitude: 0.91 μV).

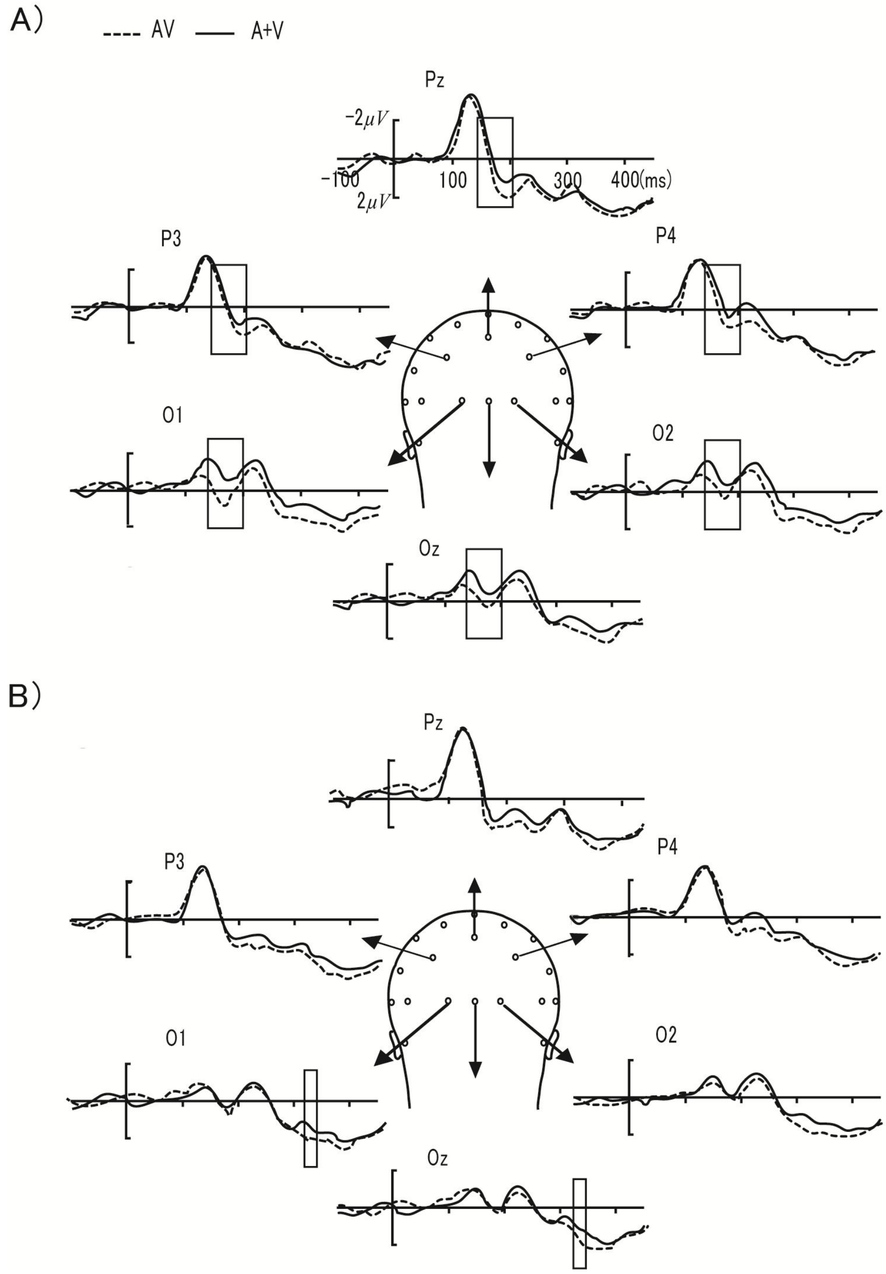

Fig.5.3 Grand-average event-related potentials elicited by audiovisual stimuli and auditory plus visual stimuli. (A) Reliable auditory stimuli condition. (B) Unreliable auditory stimuli condition. The square areas indicate the time periods when the bimodal response significantly differed from the sum of the unimodal responses.

5.4.4 Unreliable auditory stimuli's effect on Integration

In the unreliable auditory stimuli condition, no significant differences between AV and

A+V were found at 140-200ms after POS in the parietal and occipital regions around P3, Pz, P4, O1, Oz and O2 [$F(1,12) = 4.677, p = .053$] (Fig.5.3B). The difference between AV and A+V were found at 320-340ms after POS in the occipital area around at O1, Oz and O2. The analysis yielded an effect of the modality [$F(1,12) = 4.90, P < .05$]. The occipital differences between AV and A+V were confirmed at O1 [$F(1,12) = 6.08, p = .029$] (mean amplitude: 0.35 μV), Oz [$F(1,12) = 4.85, p = .047$] (mean amplitude: 0.53 μV).

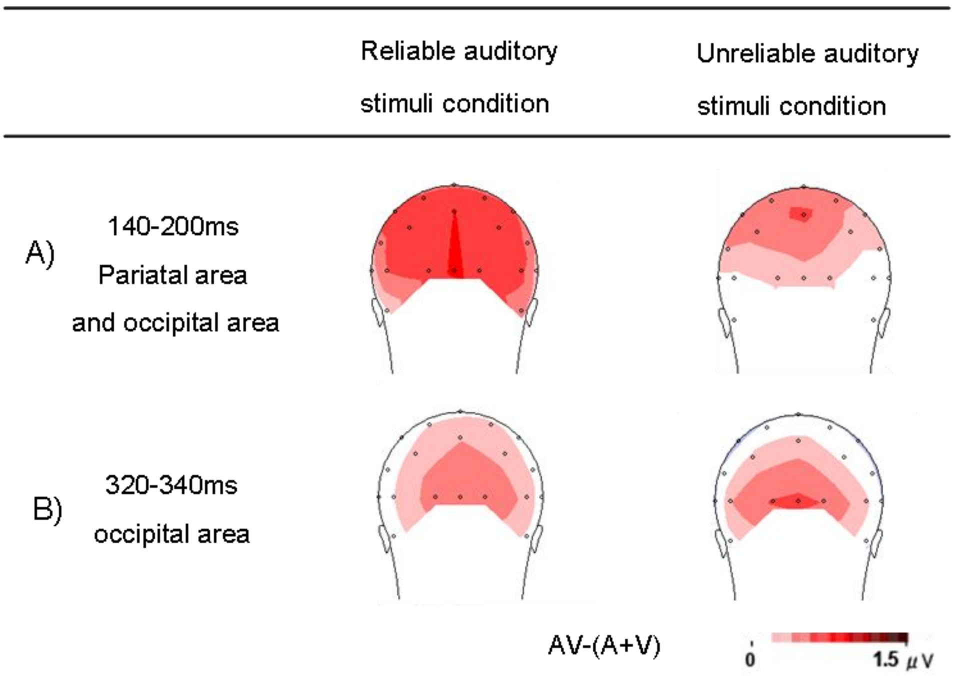

Fig.5.4 Scalp topographies of audiovisual integration. Audiovisual integration occurred (A) over the parietal and occipital regions at 140–200 ms when the auditory stimuli were reliable and (B) over the occipital region at 320-340ms when the auditory stimuli were unreliable.

5.5 Discussion

This study was deserved to reveal behavioral and electrophysiological evidences for the influence of the reliability of temporal synchrony on audiovisual integration. The reliable auditory stimulus induced audiovisual integration occurred at 140-200 ms after POS covering parietal and occipital regions. The integration effects of unreliable auditory stimulus were found at 320-340ms after POS at occipital region.

Integration effects at 140-200 ms

When the temporal information of auditory stimuli was reliable, integration effects were measured in the parietal and occipital areas between 140 and 200 ms after POS (Fig.5.4). Several previous studies of audiovisual integrations have found analogous effects at this latency around parietal and occipital areas [5.7-5.9]. The effects at posterior and occipital areas have been functionally associated with visual discrimination processes and suggest that bimodal stimuli change activities at posterior and occipital areas [3.4]. Thus, in the present study, reliable auditory stimuli affected the visual perception and presumably elicited significant integration.

The effect at 140-200 ms latency was not found in the unreliable auditory stimuli condition. A previous study by Qiangliu presented high- and low-reliability of visual stimuli and found that integration effects in the visual cortex by high-reliability visual stimuli were stronger than those of low-reliability stimuli in this latency window[3.4]. In present study, the early integration elicited by reliable auditory stimuli was found at

140-200ms. Early sensory processing was thought to be occur during the first 200ms POS [3.8], this suggests that temporal reliability impacts AV integration beginning at the early sensory processing stage.

Integration effects at 320-340ms

Unreliable auditory stimuli also elicited integration in the occipital area but this brgan later than in the reliable condition, occurring at 320-340ms. In present study, the temporal reliability of stimuli was controlled by changing probability of being presented with synchronous audiovisual stimuli. In the reliable condition, the probability of audiovisual synchrony was 100%. However, in the unreliable condition, the probability of audiovisual synchrony was 20%. In the studies of Girardi et al. and Ciaramitaro et al. , where stimuli synchrony probability differed, a high probability of synchrony could more significantly improve perception[5.10,5.11]. They found that the probability of stimuli synchrony could modulate attentional shift. This suggests that the brain might devote more attention resources to reliable stimuli for processing [5.7].Furthermore, other previous studies have observed that attention can reciprocally modulate audiovisual integration, which occurs at various stages of stimulus processing [5.12]. In a previous study by Talsma et al. in 2005, audiovisual integration occurred later when subjects were not attending to stimuli than when they were attending to the stimuli [1.9]. In the present study, the integration effects were different between the reliable and unreliable auditory stimuli conditions. It is therefore possible that the subjects were less attentive when presented with unreliable auditory stimuli; this may thus have induced later integration. Our results show clear effects of temporal reliability on audiovisual integration.

5.6 Conclusion

This study compared the differences between audiovisual integration elicited by reliable and unreliable auditory stimuli through behavioral and ERP experiments. The results show that the reliability of auditory stimuli can modulate the neural activity of parietal and occipital regions and that the integration elicited by unreliable auditory information occurs later than that elicited by reliable auditory information. The impact of temporal reliability on AV integration begins at an early sensory processing stage.

References

[5.1] Ernst MO, Bülthoff HH. Merging the senses into a robust percept. Trends in cognitive sciences 2004; 8: 162-169.

[5.2] Lippert M, Logothetis NK, Kayser C. Improvement of visual contrast detection by a simultaneous sound. Brain research 2007; 1173: 102-109.

[5.3] Li Q, Wu J, Touge T. Audiovisual interaction enhances auditory detection in late stage: an event-related potential study. Neuroreport 2010; 21: 173-178.

[5.4] Schall S, Quigley C, Onat S, König P. Visual stimulus locking of EEG is modulated by temporal congruency of auditory stimuli. Experimental brain research 2009; 198: 137-151.

[5.5] Shams L, Kamitani Y, Shimojo S. What you see is what you hear. Nature 2000.

[5.6] Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. Nature 2002; 415: 429-433.

[5.7] Liu Q, Qiu J, Chen A, Yang J, Zhang Q, Sun H-J. The effect of visual reliability on auditory-visual integration: an event-related potential study. Neuroreport 2007; 18: 1861-1865.

[5.8] Giard M, Peronnet F. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. Journal of cognitive neuroscience 1999; 11: 473-490.

[5.9] Yang W, Li Q, Ochi T, Yang J, Gao Y, Tang X *et al.* Effects of Auditory Stimuli in the Horizontal Plane on Audiovisual Integration: An Event-Related Potential Study. PLOS ONE 2013; 8: e66402.

[5.10] Girardi G, Antonucci G, Nico D. Cueing spatial attention through timing and probability. Cortex 2011.

[5.11] Ciaramitaro VM, Cameron EL, Glimcher PW. Stimulus probability directs spatial attention: an enhancement of sensitivity in humans and monkeys. Vision Research 2001; 41: 57-75.

[5.12] Koelewijn T, Bronkhorst A, Theeuwes J. Attention and the multiple stages of multisensory integration: A review of audiovisual studies. Acta psychologica 2010; 134: 372-384.

Chapter 6

General Conclusion and Future Challenges

Summary

This dissertation has investigated the human mechanism of audiovisual temporal integration by event-related potential experiments. In particular, we discussed the effects of temporal synchrony, temporal asynchrony and temporal reliability on audiovisual integration. The findings are summarized below. In addition, we describe challenges for future research.

6.1 General conclusions

The dissertation is composed of six chapters. The first and second chapters instructed the previous studies and methods of our study. The third, fourth and fifth chapters investigated the temporal characteristics of stimuli in human audiovisual integration.

Chapter 1 describes the previous studies about audiovisual integration and the effect of temporal factor. Previous studies have shown that an audiovisual stimulus with spatial and temporal coincidence can generate the largest integration effects. Many researches have shown that there is a relatively time window of integration as large as 250 ms, in which stimuli from different modalities tend to be integrated into a single unified percept. When the auditory and visual stimuli are not match in temporal characteristics, perceptual reports regarding the temporal characteristics of visual stimuli tend to be biased toward the temporal characteristics of the auditory stimuli. Temporal asynchrony of visual-auditory stimuli can influence audiovisual integration.

Chapter 2 describes the methods of this study. We used behavior and event-related potentials (ERPs) to examine human mechanism of audiovisual temporal Integration. Electroencephalography (EEG) is the recording of electrical activity along the scalp produced by the firing of neurons within the brain. An event-related potential (ERP) is the measured brain response that is the direct result of a specific sensory, cognitive, or motor event.

Chapter 3 describes the first experiment, the visual and auditory stimuli onset synchrony and only the visual stimulus was attended. This study used event-related potentials (ERPs) to demonstrate that onset synchronous task-irrelevant auditory stimuli affect the audiovisual integration. The behavioral results showed that the responses to audiovisual target stimuli were faster than that to unimodal visual target stimuli. Moreover, the ERPs were recorded in response to unimodal auditory (A), unimodal visual (V) and bimodal (AV) stimuli. Cross-modal interactions were estimated using the additive [AV - (A + V)] model. Four ERP components related to audiovisual integration were observed. (1) over ce ntral and occipital areas at around100 to 160ms; (2) over the central and occipital areas at around 160 to 200ms; (3) over the occipital areas at around 200 to 240ms. (4) over frontal-central areas at around 280to320ms. Those finding confirms the main neural activity of audiovisual integration when the visual and auditory stimuli onset synchrony.

Chapter 4 introduces the second experiment. The effects of the temporal characteristics on audiovisual integration in a visual attention task were investigated. The visual and auditory stimuli onset asynchrony (SOA= ±400 ms, ±150 ms, 0 ms), only the visual stimulus was attended. Behavioral data and Event-Related Potentials (ERPs) were recorded. The behavioral results showed that the responses to temporal asynchronous AV stimuli were more accurate than unimodal visual stimuli. When the SOA was -150ms, the reaction time was the fastest and hit rate was the highest, The ERP results showed that the N1 latency of biomodal AV was earlier than the sum of unimodal auditory and unimodal visualstimuli in auditory preceding condition. The auditory stimuli which were presented earlier than visual stimuli were regarded as cue. The temporal relationship of auditory cues to visual targets can markedly influence the accuracy and the speed of visual detection. These results suggested that the temporal asynchronous audiovisual stimuli enhanced the visual detection.

Chapter 5 compared the differences between audiovisual integration elicited by reliable versus unreliable auditory stimuli through behavioral and event-related potentials (ERPs) experiment. In the reliable auditory stimuli condition, the visual and auditory stimuli were always presented simultaneously. In the unreliable auditory stimuli condition, the auditory stimulus was randomly presented before, in synchrony with, or after the visual stimulus; the temporal information provided by the auditory stimulus was thus unreliable. The participants were instructed to focus on the visual modality and to ignore the auditory modality. We tested how the temporal reliability of the auditory stimulus affected audiovisual integration. We found that audiovisual integration occurred around 140 – 200 ms, covering the parietal and occipital regions in the reliable condition; in the unreliable condition, integration occurred later, around 320 – 340 ms, and covered the occipital region. Our results show that the reliability of auditory stimuli can modulate the neural activity of parietal and occipital regions and that the integration elicited by unreliable auditory information occurs later than that elicited by reliable auditory information. The impact of temporal reliability on AV integration begins at an early sensory processing stage.

6.2 Future challenges

The neural mechanism of audiovisual integration is very complex. We rarely pay attention to the differences in arrival time of auditory and visual inputs. In fact, the temporal factor of auditory and visual stimuli influences the processing of our brain. According to the complexity of the neural mechanisms of audiovisual integration, future studies will focus on separating the process into several single integration characteristics. For example, we will separately study the spatial integration of audiovisual stimuli, the effect of different attention, and so on.

Appendix-- Simple Introduction of EEG Apparatus

The BrainAmp MR *plus* was manufactured by BrainProduct Inc., Germany. This amplifier is a compact solution for neurophysiology research that can be combined with other units within the same product family to cover a vast range of possible application areas. This fully portable solution can be used for standard EEG/ERP recordings and can also be placed inside of the MRI bore for simultaneous EEG/fMRI acquisitions.

Thanks to its 5 kHz sampling rate per channel, the BrainAmp can be used to record EEG, EOG, and EMG signals as well as evoked potentials with a frequency up to 1 kHz. The 16-bit TTL trigger input allows the detection of a large number of markers from visual, acoustic, electrical, magnetic or other stimulation modalities. The BrainAmps can be used both with passive and active electrodes offering a great degree of flexibility.

The 32 channel units can be stacked to expand the number of channels up to 256 and combined with the BrainAmp ExG to record EEG, EOG, EMG, ECG, GSR (Galvanic Skin Response) and many other types of bipolar and auxiliary signals.

Figure A1. EEG amplifier of BrainAmp MR *plus*

Table A1. **Technical specifications of BrainAmp MR *plus***

Number of Channels per unit	32
Max. Number of channels	128

Reference Type	unipolar
MR-compatibility	Yes (for scanners up to 4 Tesla)
Bandwidth [Hz]	DC - 1000
High Pass Filter [Hz]	0.016 / 10 s AC or DC switchable
Low Pass Filter [Hz]	1000 / 250 switchable
Input Noise [µVpp]	≤ 1
Input Impedance [MΩ]	10 / 10000
Input Measurement Ground / Reference	Yes
A/D-C [bit]	16
A/D-Rate [Hz]	5000
Max. Sampling Frequency [Hz]	5000
Offset Compatibility [mV]	± 300
Operating Range [mV]	selectable: ±3.2768; ±16.384; ±327.68
Resolution [µV]	selectable: 0.1; 0.5; 10.0
CMRR [dB]	≥ 110
TTL Trigger Input [bit]	16
Synchronized Digital Trigger Input [bit]	up to 16
Max. Power Consumption [mA]	160
Power Supply	rechargeable Battery
Signal Transmission	optical
PC Interface	PCI, USB 2.0
Deblocking Function	Yes
Blocking of Unused Channels	Yes
Safety	Twin Fiber optical Transmission Protection Class II, Type BF IEC EN 60601 EMC tested, electrically safe
Classification to MDD 93/42/EEC	Class IIa
Dimensions H x W x D [mm]	68 x 160 x 187
Weight [kg]	1.1

Publications

Journal papers:

1. **Jingjing Yang**, Qi Li, Weiping Yang and Jinglong Wu: Enhancement of Visual Detection by Temporal Alignment of Visual-auditory Stimuli: A Behavioral and Event-Related Potential Study. *Journal of Information* .Vol.16, No.1(A), pp.527-534,(2013).
2. Qi Li, **Jingjing Yang**, Jinglong Wu and Noriyoshi Kakura: Spatial Location of Audiovisual Stimuli Affects the Latency of Multisensory Integration around 340 to 400 ms.*Journal of Information*, Vol.16(7)A: pp.4841-4852,(2013).
3. Qi Li, **Jingjing Yang**, Noriyoshi Kakura and Jinglong Wu: Multimodal Audiovisual Integration at Early and Late Processing Stages in Humans: An Event-Related Potential Study , *Journal of Information,* Vol.13, No.3(A), pp.807-816,(2010).
4. Weiping Yang, Qi Li, Tatsuya Ochi, **Jingjing Yang**, Yulin Gao, Xiaoyu Tang, Satoshi Takahashi, Jinglong Wu: Effects of Auditory Stimuli in the Horizontal Plane on Audiovisual Integration: An Event-Related Potential Study. *Plos one*. Vol.8, pp.1-10,(2013).
5. Xiaoyu Tang, Chunlin Li, Qi Li, Yulin Gao, Weiping Yang, **Jingjing Yang**, Soushirou Ishikawa, Jinglong Wu. Modulation of auditory stimulus processing by visual spatial or temporal cue: an event-related potentials study. Neuroscience Letters, 553(11): 40-45, (2013).

International conference papers：

1. **Jingjing Yang,** Qi Li and Jinglong Wu: Measurement and Analysis of Electroencephalogram (EEG) on Audiovisual Interaction for Early Clinical Detection of Neuropsychological Disorder, *The 3rd International Conference on BioMedical Engineering and Informatics*, pp.1877-1880.(2010.10)

2. **Jingjing Yang,** Qi Li， Yulin Gao and Jinglong Wu: Task-irrelevant Auditory Stimuli Affect Audiovisual Integration in a Visual Attention Task: Evidence from Event-related Potentials, *2011 IEEE/ICME International Conference on Complex Medical Engineering,*pp.*248-253* (2011.5)

3. **Jingjing Yang**, Tatsuya Ochi, Qi Li, Yulin Gao, Weiping Yang, Xiaoyu Tang, Satoshi Takahashi and Jinglong Wu: The Effect of Spatial Information of Auditory Stimuli on Audiovisual Interaction:Evidence from Event-related Potentials*2012 IEEE/ICME International Conference on Complex Medical Engineering.*pp.*699-704* (2012.7)

4. Qi Li, **Jingjing Yang**, Noriyoshi Kakura and Jinglong Wu: Multisensory Interaction of Audiovisual Stimuli on the Central and Peripheral Spatial Locations: A Behavioral Study, 2009 International Conferences on Brain Informatics, pp.127-134. (2009.10)

5. Qi Li, **Jingjing Yang**, Noriyoshi Kakura and Jinglong Wu: Multimodal integration elicited by audiovisual stimuli presented centrally in a divided attention task:An event-related potential study in humans, The Fifth International Conference on Information, pp.67-73 .(2009.11)

6. Yulin Gao, **Jingjing Yang,** Qi Li, Ryota Morikawa and Jinglong Wu: A Basic Study

of Event-Related Potentials (ERPs) on Human Audiovisual Spatial Integration for Human-machine Interface, *2010 IEEE International Conference on Robotics and Biomimetics.* (2010.12)

7. Qi Li, **Jingjing Yang** and Jinglong Wu: The different effects of visual attention and auditory attention on bisensory audiovisual integration, *2011 IEEE/ICME International Conference on Complex Medical Engineering.* (2011.5)
8. Qi Li, **Jingjing Yang**, Yulin Gao, Jinglong Wu: Unpredictable auditory spatial information modulates spatial-congruent audiovisual integration, The Six International Conference on Information, pp.326-331(2013).
9. Weiping Yang, **Jingjing Yang** , Takahiro Kimura and Jinglong Wu: Audiovisual integration elicited by stimuli peripherally in divided attention task: An Event-Related Potential Study, *2011 IEEE/ICME International Conference on Complex Medical Engineering.*(2011.5)
10. Siuly, Yan Li, Jinglong Wu and **Jingjing Yang**: Developing a Logistic Regression Model with Cross-Correlation for Motor Imagery Signal Recognition, *2011 IEEE/ICME International Conference on Complex Medical Engineering.* (2011.5)
11. Weiping Yang, Qi Li, Tatsuya Ochi, Jingjing Yang, Yulin Gao, Xiaoyu Tang, Satoshi Takahashi, Jinglong Wu: Influences of auditory stimuli in front and rear space on visual detection: An event-related potential study, 2013 IEEE/ICME International Conference on Complex Medical Engineering, pp. 7-12 (2013)
12. Xiaoyu Tang, C. L., Qi Li, Yulin Gao, Weiping Yang, **Jingjing Yang**, Soushirou

Ishikawa, Satoshi Takahashi, Jinglong Wu; Effect of Cue-Target Interval on Audiovisual Stimuli Processing in Endogenous Spatial Attention: An Event-Related Potentials Study, 2013 IEEE/ICME International Conference on Complex Medical Engineering, pp.656-661(2013).

Chapters in books:

1. **Jingjing Yang**,Qi Li,Yulin Gao and Jinglong Wu: Temporal Dependency of Multisensory Audiovisual Integration. Biomedical Engineering and Cognitive Neuroscience for Healthcare: Interdisciplinary Applications, Chapter 33, pp.320-326.
2. Yulin Gao, Weiping Yang, **Jingjing Yang**,Takahashi Satoshi and Jinglong Wu:Neural Mechanisms of Audiovisual Integration in Integrated Processing for Verbal Perception and Spatial Factors. Biomedical Engineering and Cognitive Neuroscience for Healthcare: Interdisciplinary Applications, Chapter 34, 2012. pp.327-336.

Acknowledgements

I express my heartfelt thanks and gratitude to my mentor, Professor Jinglong Wu, who provided valuable instruction during my study and the writing of this dissertation. I have been greatly inspired by Professor Wu's diligence, will power, courage, and kindness. His Helix Theory of Study led me to the study of cognitive neuroscience. Beyond the university, Professor Wu has also kindly helped me, a foreigner in Japan, with everyday life.

I additionally extend my great appreciation to Professor Horibe Akihiko, Tomita Eiji and Associate Professor Satoshi Takahashi, who provided valuable advice about writing the dissertation and oral presentation.

I also express my sincere thanks to Professor Tetsuo Touge of Kagawa University, who provided valuable advice about my research and kindly manuscript review.

In addition, I also thank the students and staff in Wu lab. Without their cooperation, I cannot imagine how I could have finished my experiments. I sincerely thank all those who contributed to my experiment, my paper, and my dissertation.

Finally, I extend my deep appreciation and dedicate this dissertation to my family, especially to my husband and my parents, who brought me up and have always supported and understood me.