Title of Thesis

Study on Audiovisual Interaction in Visual Detection and Discrimination by Behavioral and Event-related Potential Experiments

2018 September

Fengxia Wu

The Graduate School of

Natural Science and Technology

(Doctor's Course)

OKAYAMA UNIVERSITY

Abstract

Sensory perception consists of the detection of salient events in space and time as well as the discrimination with regard to specific features in events, and recognition. In traditionally viewing, detecting the presence of an object is a different processing than identifying the object as a particular object. However, recently, in the literature on visual categorization, researchers have reached controversial conclusions such as "as soon as you know it is there, you know what it is" or that "as soon as you know an object is there, you do not know what it is". Additionally, people obtain dynamic effective information from the complex environment through multiple senses in everyday life. Audition also is one important sensory system that human use to perceive the environment. The brain handles multiple sensory signals (80% were vision and audio) automatically and effortlessly to provide a more accurate message in order to shape and guide our behavior. Therefore, it is important to study interaction across audiovisual sensory modalities. However, the neural mechanism of audiovisual interaction is not completely clear at all. Besides, how visual stimulus (feature of spatial frequency and contrast, or visual intensity) alter audiovisual interaction is also unknown. The aim of the present study was to clarify how our brain processes audiovisual information in different perception stage and whether the processing of different visual stimulus affect audiovisual interaction.

Firstly, to clarify whether visual detection and visual discrimination depend on same mechanism, a visual detection and visual orientation discrimination task were used to test the visual threshold with different spatial frequency. Our results showed that there were no significant different in threshold between visual detection and visual discrimination, whereas the response time for visual detection were faster than that for visual discrimination. These results suggesting that the perceptual of detection and perception might rely on partially separate mechanisms.

Secondly, to investigated whether audiovisual interaction is different for different perceptual processing, a visual detection and visual identification task with/without a

task-irrelevant auditory stimulus were conducted to examine the effect of experimental task on audiovisual interaction, and the difference between different spatial frequency. The results confirmed that the response for visual discrimination was slowed, and task-irrelevant auditory stimulus speedup visual response in both visual detection and visual discrimination (so called "audiovisual interaction"), and the magnitude of audiovisual interaction were same for all spatial frequency in each task due to high contrast. However, audiovisual interaction in visual detection were larger than that for visual discrimination. Our results provided empirical evidence that complex of perceptual processing would affects audiovisual interaction.

Thirdly, to further clarify the effect of visual spatial frequency on audiovisual interaction, the visual detection task with/without a task-irrelevant auditory stimulus was performed. The results showed that spatial frequency modulates auditory facilitation of visual detection at low contrast (20%) but not at high contrast (100%). Moreover, the data revealed that audiovisual interaction was larger for low (0.54 cycles/degree) and high (6.46 cycles/degree) spatial frequencies than for a medial spatial frequency of 0.70 cycles/degree (all $p < 0.05$). However, when the visual stimulus was adjusted to the same perceived intensity for each spatial frequency by changing contrast, no significant difference was found among the different spatial frequencies ($p > 0.05$). The current results suggested that the stimulus intensity of a visual stimulus is the key factor for audiovisual interaction.

Lastly, to investigate the neural mechanism of visual intensity on audiovisual interaction, a visual orientation discrimination task with/without a task-irrelevant auditory stimulus were performed using event-related potential (ERP) method. Consisted with our previous study, behavioral results showed that task-irrelevant auditory stimulus facilitated visual discrimination, suggesting audiovisual interaction occurred. The ERP results showed that in the low intensity (3.47 c/d) condition existing the earliest integration (50-90 ms) in the left posterior region, and this audiovisual interaction was delayed from auditory cortex (50-90 ms) to visual cortex (70-90 ms), suggesting that auditory enhanced low intensity visual perception via direct or indirect connectivity from auditory cortex to visual cortex during early stage(cortico-cortical).

Moreover, the audiovisual interaction over fronto-central area were delayed with decreasing visual intensity (230-260 ms, 240-300 ms and 280-320 ms for the intensity of 1.00, 1.86 and 3.47 c/d). In addition, audiovisual interaction over parietal-occipital area were delayed with decreasing visual intensity (310-500 ms, 390-500 ms and 480-500 ms for the intensity of 1.00, 1.86 and 3.47 c/d). These results suggested that the audiovisual interaction pattern was depended on stimulus intensity, and further revealed a delayed audiovisual interaction resulting from the slowed visual processing.

According to the current situation, future studies will focus on special populations (e.g. older people, patients with headache, mild cognitive impairment, alzheimer's disease, and schizophrenia) to uncover the neural mechanism of audiovisual integration and to provide important basis for the early clinical detection and rehabilitation of special brain disease.

Table of Contents

Chapter 1 Introduction

Summary

This chapter introduces the concept of visual perceptual processing and audiovisual interaction. The previous studies of audiovisual interaction in different brain areas have also been summarized here. The technique of electroencephalogram (EEG) and event-related potential (ERP) have been introduced. At last, the purpose and contents of the thesis are briefly explained.

1.1 Visual perception

Sensory perception comprises the detection of salient events in space and time as well as their discrimination with regard to specific features or configural properties and recognition. In traditionally viewing, detecting the presence of an object is a different processing than identifying the object as a particular object. Visual perception of an object is instantly connected with an idea of what we see, but sometimes we might make a mistake and sometimes we just have the impression that there was something. Specially, in the literature on visual categorization, researchers have reached controversial conclusions such as “as soon as you know it is there, you know what it is” [1] or that “as soon as you know an object is there, you do not know what it is” [2].

Kalanit et al. (2005) have found that subjects performed just as quickly and accurately on the categorization task as they did on a task requiring only object detection: By the time subjects knew an image contained an object at all, they already knew its category. Moreover, Functional magnetic resonance imaging (fMRI) have showed that visual detection and visual identification have same visual activity over visual primary cortex. Therefore, it is reasonable that as soon as you know it is there, you know what it is.

In the other hand, fMRI studies have reported partly separated cortical mechanism for object detection and identification and the mere detection of an object is easier than its full identification [3]. The importance of this distinction is underlined by a behavioral study, showed that discrimination performance curve and aware detection curve are different relay on whether feature have been identified during perceptual [4]. Additionally, electrophysiological studies further investigated the different correlated of awareness for detection and identification, and aware detection of object’s presence has an earlier and more posterior than aware identification of the object [5]. Therefore, it is possible that as soon as you know an object is there, you do not know what it is. Thus, the problem arises from the question of where object detection end and identification start.

1.2 Audiovisual interaction

To achieve a comprehensive picture of the external world in everyday life, the brain integrates information from multiple senses. Audition and vision are two important sensory systems that humans use to perceive the environment, as 80% information was received by auditory and visual systems. The main region of the brain in which audition is perceived is the auditory cortex (AC), and the main area of the brain in which vision is perceived is the visual cortex (VC) (see figure 1.1) [6]. Although input signals about the same external environment during normal daily activities transmit through different models into the different cortical representations, these two sensory signals are automatically and effortlessly bound to provide a more accurate spatial and temporal information, in order to shape and guide our behavior. This binding process between auditory and visual signals, called audiovisual interaction.

Audiovisual interaction has been demonstrated occurred at superior temporal sulcus in both animal and human. M. A. Meredith et al. (1978) M. S. Beauchamp et al. (2005) have reported that audiovisual interaction is not merely the linear combination of two unimodal information in the superior temporal sulcus [7, 8]. Sometimes much greater than the mere sum of the individual unisensory response, related research has used the fMRI showed dramatic activity in the superior temporal sulcus (AV > A +V)[9]. They have scanned the brain activity of the subjects when responding to auditory (sounds), visual (mouth movements), or audiovisual stimuli (simultaneous sounds and mouth movements separately. And found that the brain activity AV stimulus was stronger than that for (A+V). Alternatively, related research has used the fMRI also showed dramatic activity for AV stimulus was weaker than that for (A+V) in the superior temporal sulcus (AV < A +V) [10].

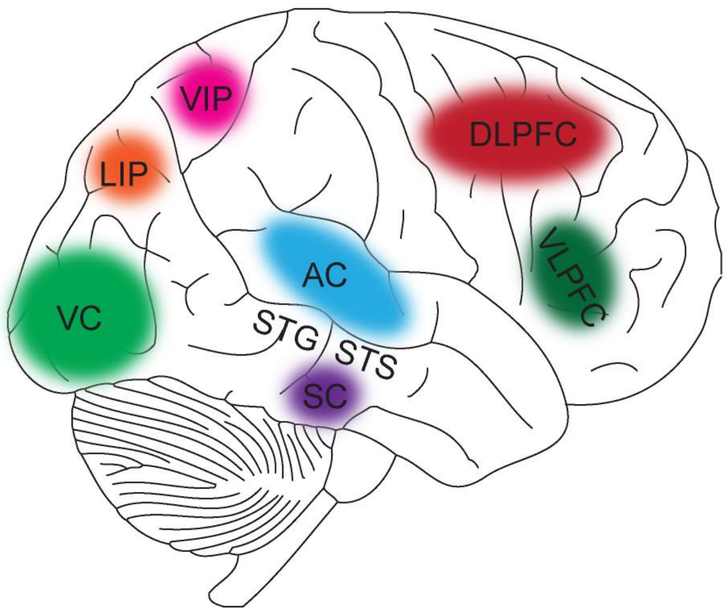


Figure 1.1 Audio-visual processing in different brain areas.

Besides superior temporal sulcus, previous studies have shown that audiovisual interaction also occurs in the auditory cortex (AC, which also receives visual inputs) [7], the visual cortex (VC, which also receives auditory inputs) [9], the ventral intraparietal area (VIP, where neurons respond to visual and auditory events), an area for three-dimensional integration [11], and a lateral intraparietal area (LIP, where have been reported respond to visual and auditory events) belonging to the visual cortex [12]. According to the investigations of multisensory integration in cat, Meredith et al. (1987, 1986) found that the superior colliculus (SC, mainly involved in eyes movement) can receive auditory and visual signals and demonstrated that the two types of signals can active the same neuron when they are presented in the same place [13, 14]. MEG studies have reported that the superior temporal gyrus (STG), a auditory processing area, also has involved in audiovisual interaction [15]. Moreover, current research have indicate that the superior temporal sulcus (STS) contains the region involved in audiovisual interaction [6, 16-18]. Senkowski et al. (2007) have examined the an super-additive BOLD activation in the STS and showed that the subjects elicited BOLD activation to audiovisual stimuli greater than the sum of BOLD activation to individual auditory or visual stimuli [19]. Barraclough et al. studied audiovisual integration by examining the neural level of non-human primates, and they found that the sound of actions modified 23% of the visual response of the STS neurons coding the sight of actions [18]. In addition, Dorsal Lateral Prefrontal Cortex (DLPFC) which is in charging of attention, have been found response to audiovisual stimuli [20, 21]. Ventral Lateral Prefrontal Cortex (VLPFC), which is in charging of working memory, also have been indicated respond to audiovisual stimuli [22] (see Figure 1.1).

1.3 Event-related potentials (ERPs)

1.3.1 Event-related potentials (ERPs)

ERPs are measured by means of electroencephalography (EEG), EEG recordings show the overall activity of the millions of neurons in the brain. The recording shows fluctuations with time that are often rhythmic in the sense that they alternate regularly. The EEG patterns change when external stimuli (such as sounds or pictures) are presented, whereas ERPs is the measured brain response that is the direct result of a specific sensory, cognitive, or motor event by non-invasive method. The transient electric potential shifts (so-called ERP components) are time-locked to the stimulus onset with the present trigger to marking the onset time (Figure 1.3). Each component reflects brain activation associated with one or more mental operations. Contrasting with behavioral measures such as response times, ERPs are characterized by simultaneous multi-dimensional online measures of polarity (negative or positive potentials), amplitude, latency, and scalp distribution. Therefore, ERPs can be used to identify and distinguish neural and psychological sub-processes involved in perceptual, motor, or cognitive tasks.

Figure 1.3 Schematic of the ERP data

1.3.2 Analysis method of ERPs data in the audiovisual interaction study

The ERPs elicited by the task-irrelative stimuli were analyzed. The data were band-pass filtered from 0.01 - 60 Hz during recording at a sample rate of 500 Hz. The data were divided into epochs, from 100 ms before to 600 ms after the stimulus onset, and baseline corrections were made against a 100 ms to 0 ms time interval before stimuli onset. Trials with a voltage exceeding ± 100 mV relative to baseline were rejected automatically from the analysis. In addition, the data associated with a false alarm were excluded. The data were then averaged for each stimulus type, following digital filtering with a band-pass filter of 0.1 - 30 Hz, and the grand-averaged data were obtained across all participants for each stimulus type (V, A and AV) in each electrode. The previous studies showed that audiovisual integration was assessed by the difference wave [AV - (A+V)], obtained by subtracting the sum of the ERP waves of the unimodal stimuli from the ERP waves of the bimodal stimuli [23, 24], and the logic of this additive model is that the ERPs to bimodal (AV) stimuli are equal to the sum of the ERPs to the unimodal (A+V) stimuli, plus the putative neural activities specifically related to the bimodal nature of the stimuli. If there is significant difference between AV and (A+V), the interaction between vision and auditory is occurred (Figure 1.4).

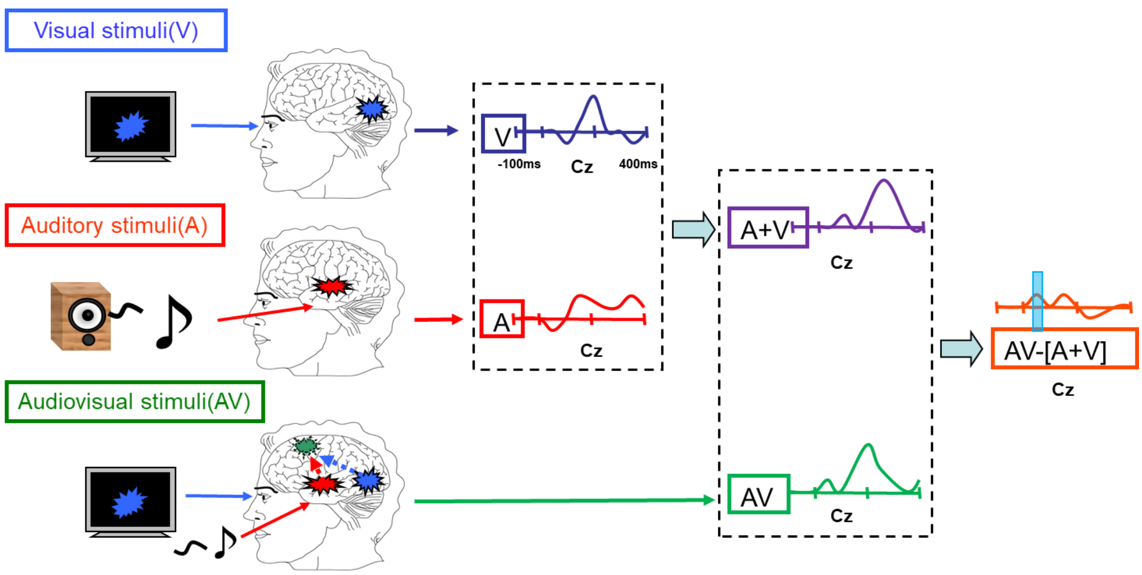


Figure 1.4 Analysis methods of ERPs data in the audiovisual integration study

1.4 The purpose of the present dissertation

The aim of this thesis studies was to investigate the brain activities of cross-modal audiovisual interaction using behavioral and electroencephalography (EEG) with high temporal resolution and to elucidate the mechanism of audiovisual interaction in different perceptual processing stage.

1.5 The contents of the dissertation

This dissertation mainly investigates brain mechanisms of audiovisual interaction with visual detection and visual discrimination tasks. In addition, further clarify the effect of visual feature on audiovisual interaction with vary spatial frequency.

Chapter 1 introduces the concept of audiovisual interaction in the brain, related previous studies, Electroencephalogram (EEG) analysis, and the method of event-related potential (ERP) analysis in audiovisual interaction studies. The aim and contents of the thesis are also briefly described.

Chapter 2 describes the first experiment. To clarify whether visual detection and visual discrimination depend on same mechanism. The present study investigates the perceptual of detection and perception might rely on partially separate mechanisms.

Chapter 3 describes the second experiment. The effects of perceptual complex on audiovisual interaction. Visual detection and visual discrimination task in audio-visual environment were designed. Audiovisual interaction was compared the between visual detection and visual discrimination using auditory facilitation effect.

Chapter 4 describes the third experiment. In this experiment, audiovisual interaction elicited by vary spatial frequency in visual detection task was investigated, auditory facilitated visual detection depended on visual intensity.

Chapter 5 describes the fourth experiment. In this experiment, audiovisual integration elicited by stimuli intensity was investigated using behavioral and electrophysiological measurements in visual discrimination task.

Chapter 6 provides a general conclusion based on the findings of the four experiments and future challenges.

Chapter 2 Spatial frequency processing in visual detection and discrimination task

Summary

Although previous studies have demonstrated visual processing in detection and discrimination task in human, how detection and discrimination task alter visual processing has not yet been completely elucidated. To investigate this issue, we design performing two classical perceptual tasks: grating detection and grating orientation discrimination. In grating detection, participants were instructed to identify stimulus, if they saw a stimulus, by pressing right button as quickly and accurately as possible. In grating orientation discrimination, participants were told to identify the orientation of the stimulus and pressed relevant button as quickly and accurately as possible. We assessed the diversity by measuring the magnitude of sensitivity and intercepts through reaction times (RT). The results showed that RT strongly depended on experimental task. The response to discrimination task is significantly slower than that for detection task ($p < 0.05$). However, visual detectability was not depended on experimental task, and no significant difference of sensitivity was found among detection task and discrimination task ($p > 0.05$). Our results provide unique insight into how the brain processes visual signal of different experimental task.

2.1 Background

Human visual system is more sensitive to contrast than absolute luminance and perceive the world similarly regardless of the huge changes in illumination over the day or from place to place. Contrast has long been known to lead to visual changes. The spatial frequency is a characteristic of any structure that is periodic across position in space, it also been known related to visual changes. Previous studies have showed that orientation have effect on contrast detection [25]. Neurons in primary visual cortex of cats and monkeys respond selectively to the orientation of grating stimulus, showed that neurons are capable of signaling orientation differences [26]. However, these experiments were restricted to high contrast grating stimulus. For low visual contrast, it is suggested the orientation didn't alter detectability when the changed of orientation was less than 10 degree [27]. In addition, some studies have found that primary visual cortex is required for detection and discrimination of visual features [28, 29], whereas others argue that primary visual cortex is required for discrimination but not detection [30], and yet others identified only subtle changes in visual acuity [31, 32]. Human behavioral studies also reported that detection and discrimination have different effect on visual grating processing [33]. These studies suggested that visual processing was related to experimental task. However, it is not completed clear whether and how experimental task influence the visual processing in human. The aim of present study is to investigate how experimental task influence human visual contrast and spatial frequency (SF) processing. Visual detection task and visual discrimination task were used to measure visual processing. RT to visual stimuli is determined by contrast level and SF [34]. The decreasing of RT to increments in contrast is well accounted for by the Pieron function in *Eq. 1* [35, 36].

$$\mathrm{RT} = k \times c^{-\alpha} + t_0 \tag{2.1}$$

In this case, α and k modulate the decay of the RTs caused by stimulus-dependent variables, in many cases it is possible to assuming an α exponent of -1, which is a particular case of the general function often applied in visual psychophysics. The slope k can be interpreted as the gain rate for the presence stimulus (1/ k is meaning for sensitivity). t_0 represents the asymptotic RT, which reflects a time constant that includes processing latencies intrinsic of sensory pathway and motor time of the effector system. According to the visual literature, k has been shown to modulated by the different contrast and SF [36]. In present study, we used this framework to investigate how experimental task modulates processing of visual stimulus.

2.2 Methods

2.2.1 Participants

Five students from Okayama University take part in this experiment, and the age is range from 22 to 24 years (mean age 23.2 years). Participants had normal or corrected-to-normal vision and normal hearing capabilities and right-handed, and they have provided written informed consent for their participation in this experiment, which was previously approved by the ethics committee of Okayama University.

2.2.2 Stimuli

Stimulus was a rectified Gabor patches with vertical sinusoidal grating, the size of visual stimulus was approximately 5° diameter (43.7 mm) and presented at center. There were seven kinds of Gabor patches was SF of 1.00 and 6.46 cycles/degree with different contrast which was range from 1% to 4% and from 3% to 20% in the units of Michelson contrast ((max - min)/ (mix + min)), respectively. The max and min being maximal and minimal value of the Gabor patch, and different contrast values were equally likely. Two fixation point were black circles with 2.2 × 2.2 mm, presented at below and upper 5° from the center.

2.2.3 Procedure

The experiment was performed in a dark, sound-attenuated and electrically shielded room (laboratory room, Okayama University, Japan). Participants sat on a comfortable chair with their head fixed by a chin-rest. Each participant completed eight blocks with two different tasks. The two tasks were measured in a random order as generated a by randomizing function.

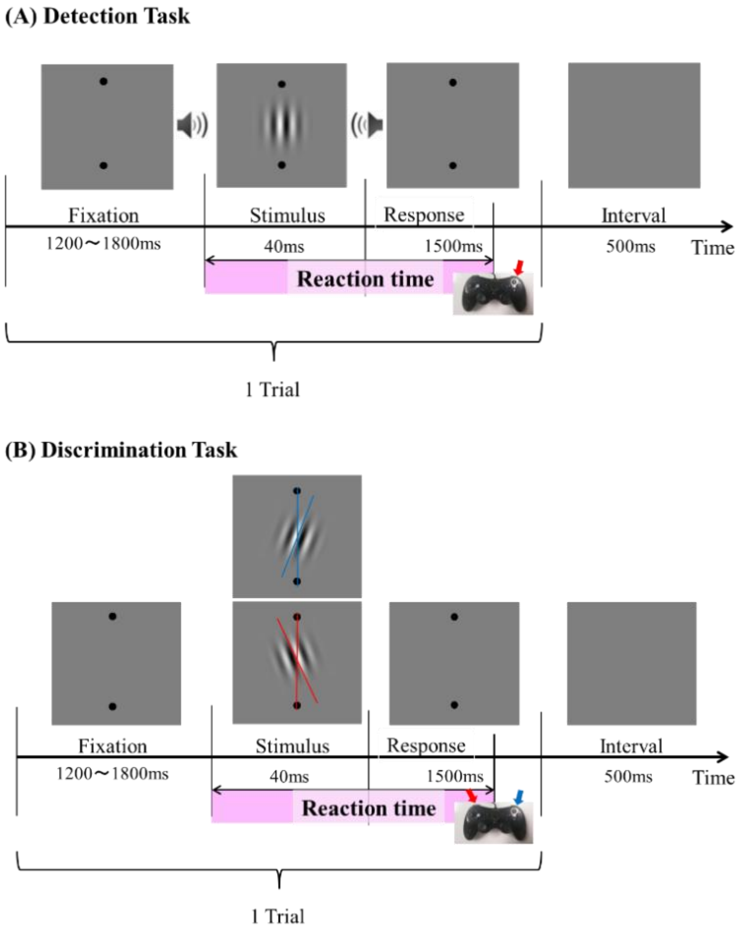


Figure 2.1 Experimental design. (A) Schematic representation of detection task for one trial sequence. Subjects sat approximately 70 cm from the screen. The subjects keep eyes on center. Stimulus were presented in the front screen and presented 40 ms. There were seven kinds of visual contrast range from 0.8 % to 4%. The participant's task was to make a

speeded button (right) response when stimuli were presented as quickly and accurately as possible. (B) Schematic representation of Discrimination task for one trial sequence. The participant's task was to identify the orientation of stimulus were deflection on the right (by pressing the right button) or the left (by pressing the left button) as quickly and accurately as possible.

Each block consisted of 70 stimuli, all the stimuli were presented randomly, and each stimulus was presented 40 ms, and the inter-stimulus interval ranged from 1 200 ms to 1 800 ms. At the begin, it would duration 1 750ms before the first stimuli appeared. They were asked to maintain pay attention to the fixation point (center from two points). Practice blocks were run as longer as participants discriminate the target stimuli and understand the task, usually about 3 min. Regardless of the subject had responded to the stimulus or not, program will continue with the next trial at the set ISI time (1 500ms). Participants were given three minutes to relax after each block to make data accurately.

Detection Task (Figure 2.1 A): there were four blocks for detection task. In these blocks, the orientation of visual stimulus was always vertical, the participants were instructed to identify stimulus, if they saw a stimulus, pressed right button as quickly and accurately as possible. Discrimination Task (Figure 2.1 B): there were also four blocks. In these blocks, there are two orientations of +10° or -10° defected from vertical. The participants were told to identify the orientation of the stimulus. When stimulus was deflected 10° to right, by pressing the right button as quickly and accurately as possible; and when stimulus was deflected 10° to left, pressing the left button as quickly and accurately as possible.

2.2.4 Apparatus

The stimuli were generated and controlled using MTLAB with Psychophysics

Toolbox for Windows 7 and displayed on a revised linearized CRT (100 Hz,1 024 × 1 024, mean luminance 20 cd/m2 and maximum luminance 116 cd/m2) situated 70 cm in the front of the participant's eyes.

2.2.5 Data analysis

Hit rate and RTs for different condition were computed. Hit rate was the number of correct reflects to target stimuli divided by the total number of target stimuli. At first, the data of RT were calculated by subject's response time for the correct responses stimuli. These results were analyzed using a repeated-measures analysis of variance with the task type (detection and discrimination) as subject factors. Secondly, we tested the Pieron function by Eq. 1, and curves were fitted to the data using a maximum likelihood estimate of k and t_0. A 2 Task type (Det., Dis.) × 2 SF (1.00, 6.46) repeated-measures analysis of variance with the parameters of k and t_0. The level of significant was fixed s at corrected $p < 0.05$.

2.3 Results

RTs and Hit rate: A 2 Task type (Det., Dis.) ×8 Contrast level (C1 – C8) ANOVA for hit rates of 1.00 were statistically expressed as no main effect of the factor of task type, [$F(1, 4) = 4.84$; $p = 0.093$], showed that the detectability was not changed by task. The main effect for contrast was significant, [$F(7, 28) = 201.79$; $p < 0.001$], showed the detectability was increased with contrast. A Task type (Det., Dis.) ×8 Contrast level (C1 - C8) mixed model ANOVA analyzed the RT for 1.00 to stimuli showed a main effect of the factor of task type, [$F(1, 79) = 276.00$; $p < 0.001$], suggested that response times to detection task was faster than that for discrimination task. The main effect of contrast also significant, [$F(1, 79) = 609.73$; $p < 0.001$], showed that responses times to high contrast was significant faster than that to low contrast in both detection task and discrimination task, see Figure 2.2. As there were similarly results for spatial frequency of 6.46, we will further test the diversity by *Perion function*.

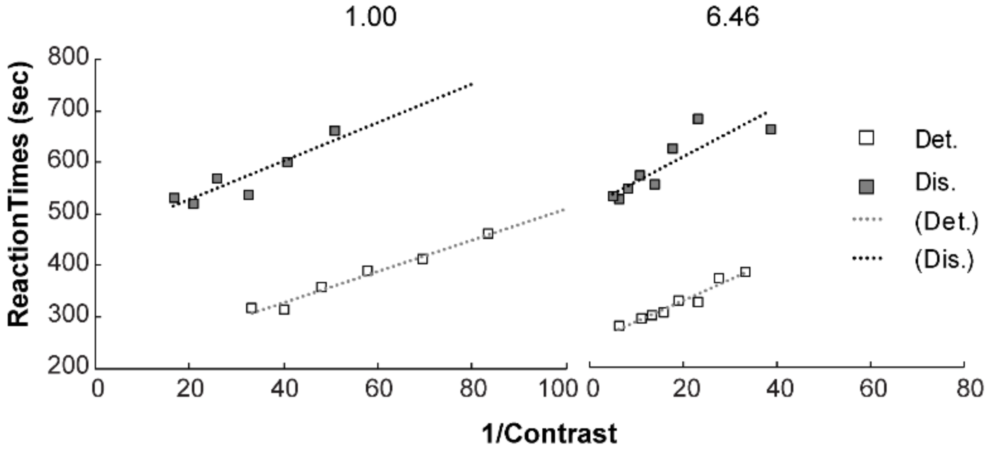


Figure 2.2. Response times as a function of the reciprocal of contrast (1/C) for each SF (from left to right) and experimental task combination (plotted in different shades of gray) in one subject. Small symbols represent the means of each condition. Lines represent the linear least-squares regression fits for experimental task. It can be appreciated that that the slope of these fits (k parameter) is shallower in the high SF condition.

K and t_0: We adjusted the RT data as a linear function of the reciprocal of contrast for each different spatial frequency and experimental task combination (Figure 2.3 or Table 1) to estimate the corresponding slops and t_0. The mean value of sensitivity (*1/k*) and t_0 for visual detection and discrimination was showed in Table 1. To test the effect of experimental task on visual sensitivity, a 2 Task type (Det., Dis.) × 2 SF (1.00, 6.46) repeated-measures analysis of variance on sensitivity was expressed a significant main effect for SF, [$F(1,4) = 8.57$, $p < 0.043$], suggest that sensitivity for SF of 1.00 was significant faster than that for 6.46 (489 vs. 165). The results were in agreement with typical contrast curves. Further post-hoc tested showed that the difference of SF was significant at detection task and there was a trend for difference in discrimination task,

p was 0.033 and 0.083, respectively (see Fig. 2.3 A). However, no significant difference was found in experimental task and there was no interaction between experimental task and SF. Our data suggesting that experimental task not influence visual detectability.

Table 1. Mean data over all participants in the parameter k and t_0.

SF	Sensitivity (1/k)		t0 (ms)	
	Det.	Dis.	Det.	Dis.
1.00	546 (92)	432 (137)	233 (09)	437 (22)
6.46	201 (41)	128 (27)	272 (20)	502 (26)

Standard error of the mean (SEM) is given in parentheses. Det. means detection task and Dis. means discrimination task.

We ran another ANOVA on the intercepts (t_0) extracted from the fits to *Eq. 1*, using the same 2 Task type (Det., Dis.) × 2 SF (1.00, 6.46) repeated-measures analysis of variance. This analysis revealed a significant main effect of Task type, [F (1,4) = 82.03, p < 0.001], indicating an overall RT increased in discrimination task with respect to detection task (a speed down of 216 ms on average). The post-hoc comparisons showed that t_0 to detection task was fast than that to discrimination task for both spatial frequency of 1.00 and 6.46. This enhancement was equivalent for all conditions (see Fig. 2.3 B). In addition, there were revealed by lack of significant interaction between Task type and SF, [F (1,4) = 9.68, p = 0.62]. These results suggesting that experimental task influence processing latencies intrinsic of sensory pathway and motor time of the effector system.

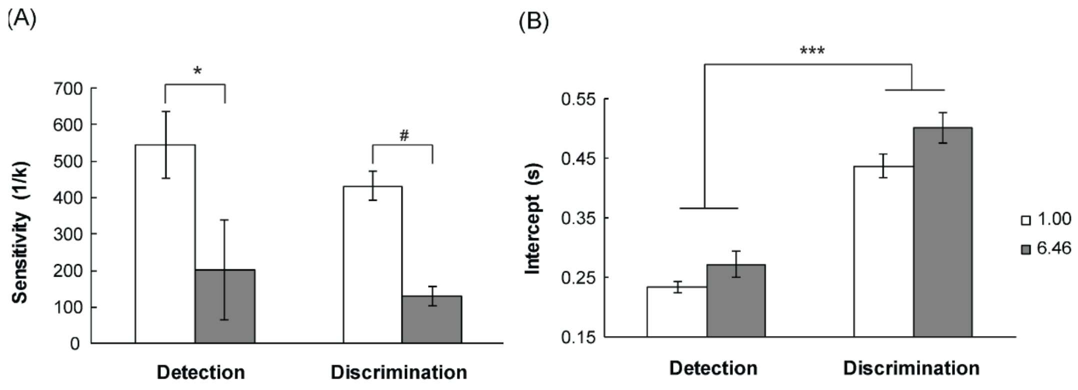


*Figure 2.3 The results of sensitivity (A) for Detection task and Discrimination task in spatial frequency of 1.00 and 6.46. (B) intercepts for Detection task and Discrimination task in spatial frequency of 1.00 and 6.46, # $p = 0.08$, * $p < 0.05$, *** $p < 0.001$.*

2.4. Discussion

In the present study, we explored the diversity of detection and discrimination on visual processing by behavioral methods. To investigate this issue, visual detection task and visual discrimination task were performed in this study. In the detection experiment, subjects were instructed to do a detection task, when they saw visual stimulus by pressing right button. In the discrimination experiment, subjects were instructed to do a discrimination of visual orientation, when stimulus deflected to right pressed right button, when stimulus deflected to left pressed left button.

Our results showed that no significant difference for hit rates between detection task and discrimination task. These results suggested that correct perceptive is basic detectability, not changed by experimental task, which was consisted with previous studies, which showed that hit rate is increased with contrast increasing [37-40]. Additionally, analysis for sensitivity also showed no significant difference between visual detection task and visual discrimination task. Therefore, our results suggested that experimental task not altered basic visual detectability. We claim that the changes in the slope are linked to sensory processing independently of the model applied and are specifically larger for the low-SF visual channels than for the high-SF channels, highlighting their stimulus dependence. However, some previous studies have showed that discrimination is better than detection [39, 41, 42]. In the study of Dzhafarov et al. (1982), they did visual position perceive, and reported that the performance for visual discrimination will better than for visual detection in RTs but not hit rates. Therefore, it is reasonable that experimental task not influenced the basic visual detectability.

The results described above are clear in that response times to visual events are modulates by experimental task, RTs for detection were significant faster than that for discrimination. The results were agreement with previous studies, which reported that task influenced reaction times [34, 43-47], due to the complexity of the task [38, 43, 44, 48-51]. In the study of Sagi et al. (1984), visual detection and visual discrimination task was preformed, showed that the response time for visual detection was significantly faster than that for visual discrimination (130 ms vs. 180 ms). Recent studies also reported the relationship between visual search and visual detection [52], showed that the speed of search was slower than visual detection and search effectiveness was depended on the size (complexity). Gilbert et al. (2013) have argued that the speed of visual processing was depended on the task [44]. In addition, RTs is known to decreased with increasing contrast [35, 45, 53]. In the study of Brooks et al (2001), a visual detection task with vary contrast were preformed, they reported that visual RTs for high contrast was significant faster than that for low contrast. It would be very surprising if the same exponent applied to contrast and any other measure of stimulus strength. Some other studies have measures of stimulus strength, found that RTs was increased with increasing spatial frequency [35, 54]. Therefore, our results provide some evidences that the speed of stimulus processing is related to the experimental task and stimulus strength.

Additionally, we have compared t0 for detection and discrimination thresholds, there was also diversity between detection and discrimination task (0.253 s vs. 0.469 s). Experimental task not alter basic visual perceive by the means of early sensory processing. Although it is rare that we view low contrast stimuli in otherwise visual environments, considerable emphasis has been placed on psychophysically determined contrast under these conditions. Therefore, we considered that the experimental task altered the speed in latencies intrinsic of sensory pathway and motor time of the effector system. Additionally, we proved some basis data from framework for robot processing system, see Figure 5. We compared the t0 for spatial frequency of 1.00 and 6.46, but no significant diversity was found. This result is not consisted with RTs, it may cause by the number of participants. Therefore, our results do not allow us to draw conclusion about the visual strength of spatial frequency influence visual processing (even many previous studies had proofed this). Some studies have argued that visual detectability in early visual processing [36-38], which caused by basic visual property binding effect. However, the influence of experimental task on this binding effect is needed further to confirm.

2.5. Conclusions

The present study suggested that visual detectability is not depended on experimental task, revealing no difference was found among detection task and discrimination task. However, the RTs is strongly depended on experimental task, showing response times to discrimination task is significantly slower than that for detection task, which based on latencies intrinsic of sensory pathway and motor time of the effector system.

Chapter 3 Visual discrimination task attenuates audiovisual interaction regardless of spatial frequency

Summary

Although previous studies have shown that task-irrelevant auditory stimuli can facilitate visual perception, it remains unclear whether this audiovisual benefit in detection and identification processes can be attribute to the same mechanism. To clarify this, we instructed participants to perform a visual detection task and a visual identification task with/without a task-irrelevant auditory stimulus in the present study. Our results showed that the task-irrelevant auditory stimulus quickened both visual detection (11.61%) and visual identification (6.11%) in all conditions. Moreover, the extent of the auditory quickening was influenced by the task demands (detection and identification) ($p < 0.001$) but was not mediated by spatial frequencies ($p = 0.533$). In addition, no interaction was found between the task demands and spatial frequencies ($p = 0.939$), indicating that the spatial frequencies and task demands influenced the audiovisual interaction independently. These findings suggested that detection and identification, modulated by audiovisual interaction rely on partially different mechanisms.

3.1 Background

In daily life, our brains handle multisensory information in an extremely efficient way and improve behavior, as seen in more rapid and more accurate responses [23, 55]. For example, when communicating with others, a speaker will generate sounds that reach our ears after the corresponding visual signals have reached our eyes, making the speech easier understand when looking at the speaker's lips. This phenomenon is known as multisensory integration [56].

Since initially introduced by Stein et al. [57], sound-induced visual improvement has been widely investigated in recent studies as a case of audiovisual interaction [55, 58, 59]. Some researchers have proposed that sound-induced improvement of visual detection originates from both the perceptual stage of processing [55, 60] and response bias (c) [58]. A typical study that supported this argument was conducted by McDonald et al. (2000) in which visual improvement was evaluated using a signal detection measure. The researchers found that the presentation of a task-irrelevant sound facilitated subsequent light detection by increasing both perceptual (sensitivity, d') and decisional measures (response bias, c). Moreover, Li et al. (2015) also used a signal detection measure to evaluate the effect of task-irrelevant auditory stimuli on visual orientation identification [61] . Similar to visual detection, the same result was found for visual identification. In particular, Chen et al. (2011) used a visual detection task and a visual orientation identification task to more clearly understand the cross-modal facilitation effect in one experiment. Their results revealed that performance of both visual detection and visual orientation identification were enhanced by the presentation of a specific (22 dB) and simultaneous noise [62]. Therefore, it is probable that visual detection and visual identification modulate audiovisual interaction by relying on the same mechanism.

On the other hand, other researchers have proposed a dissociation between audiovisual interaction in detection and identification tasks [63, 64]. That is, audiovisual interaction in detection and identification tasks may originate from different perceptual processes. For instance, Cecere Roberto et al. (2014) performed line orientation identification and visual detection studies in a patient with bilateral occipital lesions that spared residual portions of the V1/V2 area [63]. Their results showed that looming sound selectively enhanced the patient’s line orientation identification sensitivity (d’) in his relatively intact visual field, but visual detection was enhanced both in the intact and blind field, suggesting that audiovisual interaction during visual detection and line orientation identification might depend on different areas involved in perceptual processing. In addition, Gleiss Stephanie et al. (2013) examined whether and how sound enhances visual detection or visual identification performance [64], and they found a statistically significant perceptual enhancement with congruent sound at peripheral locations for visual detection only and not for identification. Recently, Kayser et al. (2017) further investigated these audiovisual congruency-facilitated perceptual benefits by using EEG. Their results revealed that sound facilitated visual motion discrimination in late stages of processing (approximately 350 ms), thereby providing the strongest evidence for the differences in response bias [65]. Therefore, detection and identification modulated audiovisual interaction may rely on partially separate mechanisms. However, as outlined above, the distinction between visual detection and identification in audiovisual interaction remains controversial.

In the present study, we focused on one specific facet of audiovisual interactions, namely, the magnitude of auditory speedup visual perception, to assess audiovisual interaction between visual detection and visual identification directly. In the visual detection task, participants were instructed to respond to all visual spatial frequency stimuli while ignoring the auditory stimuli. In the visual identification task, participants were instructed to choose one of three visual spatial frequencies to respond to while ignoring the auditory stimuli. In the contrastive analysis, the difference in the reaction times to the visual stimuli and audiovisual stimuli was defined as the audiovisual interaction. Moreover, for unimodal visual sensory stimuli, detecting the presence of an object involves a different process than identifying the object as a particular object [66]. Therefore, we expected that audiovisual interaction could be dissociated depending on whether the spatial frequency is processed during visual perception.

3.2 Methods

3.2.1 Participants

Sixteen volunteers (age range, 22-29 years; mean, 24.4 years) participated in this experiment. All the participants had normal or corrected-to-normal vision and audition and were right-handed. 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

The visual stimuli were displayed on a linearized 17-in CRT monitor (100 Hz, 1 280×1 024, mean luminance=10 cd/m2 and maximum luminance=65 cd/m2) positioned 70 cm from each participant's head (see Figure 3.1). The visual stimulus (V) was a Gabor patch with vertical gratings (4×4 cm, subtending approximately 2 degrees), with three spatial frequencies of 1.00, 3.47 and 6.46 cycles/degree, and presented approximately 4° below the fixation point. To avoid the intrinsic properties of the visual system during the two tasks, 100% contrast was used in the experiment. The visual stimulus was presented for 40 ms. The auditory stimulus (A) was a 3 000-Hz, 65-dB pure tone. The auditory stimulus was presented for 40 ms through headphones (MDR-1RNC, Sony, Japan) with a linear rise and fall time of 5 milliseconds. The audiovisual stimulus (AV) consisted of a visual stimulus and an auditory stimulus, in which the auditory stimulus was simultaneously accompanied by visual stimuli of varying spatial frequencies and was also presented for 40 ms.

3.2.3 Procedure

The study took place in a dimly lit, sound-attenuated room (laboratory room, Okayama University, Japan), and the experiment was generated by a custom-made program written in MATLAB with the Psychophysics Toolbox [67]. Participants sat on a comfortable chair with their heads fixed by a chin-rest. Each participant completed two tasks: a visual detection task and a visual identification task, and the order of the task presentation was counterbalanced across subjects.

In the visual detection task, each subject completed five sessions, with each session lasting approximately 6 min and there being 50 trials of each stimulus type. At the beginning of each session, subjects were presented with a fixation cross for 3 000 ms. Following fixation, there was an inter-stimulus interval (ISI) that varied from 1 200 to 1 800 ms randomly. After the ISI, the stimulus (visual, auditory, audiovisual) was presented randomly, and the subjects were instructed to identify whether a visual stimulus was presented. They were instructed to press the left button if they detected visual stimuli (see Figure 3.1).

In the visual identification task, the setup of the session was similar to that of the visual detection session; however, a unimodal auditory stimulus was not presented. Participants were asked to indicate the presence of the target stimulus, which was provided in the instructions at the beginning of each session, and to withhold a response if the target stimulus was absent. For example, when subjects were asked to respond to a spatial frequency of 1.00, they needed to respond to a spatial frequency of 1.00, and not respond to a spatial frequency of 3.47 or 6.46, as quickly and accurately as possible by pressing the left button of the mouse, regardless of whether an auditory stimulus was presented. In this task, each subject completed six sessions; each kind of target stimulus contained two sessions, and each session lasted approximately 6 min, with 25 trials for each stimulus type.

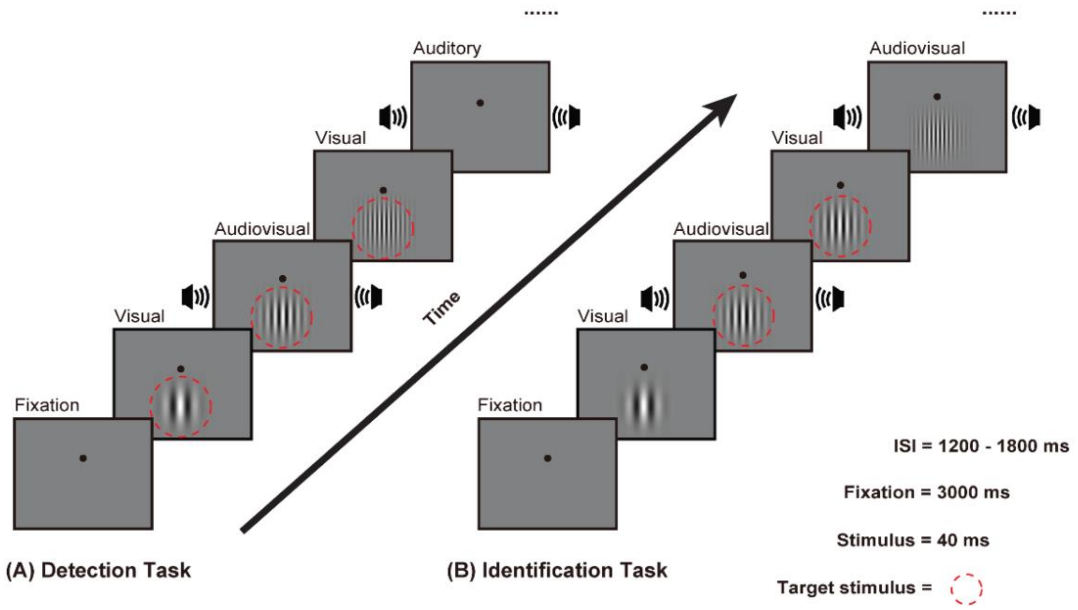


Figure 3.1 Experimental design. (A) Visual detection task. (B) Visual identification task.

3.2.4 Data analysis

Hit rate was defined as the number of correct responses divided by the total number of target stimuli. Reaction times (RTs) referred to the time of correct responses between the onset of the target and the motor response, and RTs less than 200 ms and more than 1 000 ms were excluded from the analysis. The differences in hit rates and RTs were analysed using a 2 task demands (detection, identification) × 2 stimulus types (V, AV) × 3 spatial frequencies (1.00, 347, 6.46) repeated-measures analysis of variance

(ANOVA). The difference in the RTs to the V and AV stimuli was defined as the audiovisual interaction, assessed through the interactive index [68]. Differences in audiovisual interactions within the participants were analysed using a 2 task demands (detection, identification) × 3 spatial frequencies (1.00, 3.47, 6.46) repeated-measures ANOVA. The Greenhouse-Geisser epsilon correction was used for non-sphericity, and the level of significance was fixed at a corrected $p < 0.05$.

$$\text{Audiovisual interaction}=\frac{max(\mathrm{A;V})-AV}{\max(A;V)}\times 100\% \quad (3.1)$$

3.3 Results

Hit Rate: The overall hit rates were greater than 90%, as shown in Table 1. A 2 task demands × 2 stimulus types × 3 spatial frequencies repeated-measures ANOVA of the hit rates showed a main effect of spatial frequencies only, $F (2, 30) = 7.49$, $p = 0.002$, $\eta p2 = 0.333$. Further pairwise comparisons showed that the hit rates for 1.00 were significantly larger than 6.46. However, no other significant differences were found (all $p > 0.05$).

Table 3.1. Mean RTs and hit rates in all conditions of the detection and the identification.

Stimulus Type	Detection		Identification	
	RTs (ms) ***	Hit rate (%)	RTs (ms) ***	Hit rate (%)
V 1.00	304±9.71	98.3±0.77	403±12.53	98.5±0.50
V 3.47	325±10.07	95.8±1.68	488±14.52	97.0±0.60
V 6.46	342±10.36	92.5±1.25	448±11.90	97.0±0.79
AV 1.00	269±8.75	98.4±0.49	375±12.13	99.1±0.30
AV 3.47	286±7.79	97.3±0.73	485±14.81	96.9±0.75
AV 6.46	303±8.86	95.8±1.34	423±12.58	97.9±0.83

*Data are presented as the mean ± standard error of the mean (SEM). ***p <0.001 indicates a statistically significant difference between the V stimuli and the AV stimuli in all conditions.*

Reaction Times: The mean RTs for the V and AV stimuli are presented in Table 1 and Figure 3.2. The 2 task demands × 2 stimulus types × 3 spatial frequencies repeated-measures ANOVA of the RTs revealed a main effect of the task demands, $F(1, 15) = 297.74$, $p < 0.001$, $\eta p2 = 0.952$, showing that the RTs during the detection task were faster than those during the identification task. The main effect of the stimulus types was also significant, $F(1, 15) = 177.77$, $p < 0.001$, $\eta p2 = 0.922$, indicating a faster response to the AV stimuli than to V stimuli. Importantly, the interaction between the task demands and stimulus types was significant, $F(1, 15) = 8.69$, $p < 0.01$, $\eta p2 = 0.367$, indicating that the auditory stimuli had differential effects on visual detection and visual identification. The post hoc comparisons showed that the response times to the AV stimuli were significantly faster than those to V stimuli for all spatial frequencies in both the detection task (all $p < 0.001$) and the identification task (all $p < 0.001$). However, the auditory effects seemed to be larger during the detection task (39 ms) than during the identification task (27 ms). Additionally, the main effect of spatial frequencies was also significant, $F(2, 30) = 81.70$, $p < 0.001$, $\eta p2 = 0.845$, indicating that the response times were slower from 1.00 to 6.46 ($1.00 < 3.47 < 6.46$, all $p < 0.05$). There was also a significant interaction between the task demands and spatial frequencies, $F(2, 30) = 32.98$, $p < 0.001$, $\eta p2 = 0.687$, indicating that each task had a different effect on visual spatial frequency processing. The post hoc comparisons found that the response times to the 1.00 frequency were faster than those to the 3.47 ($p < 0.001$) or 6.46 ($p < 0.001$) frequencies. Additionally, the response times were faster for 3.47 than for 6.46 ($p < 0.001$) in the detection task (RTs: $1.00 < 3.47 < 6.46$), but the response times for 3.47 were slower than those for 6.46 ($p < 0.01$) in the identification task (RTs: $1.00 < 6.46 < 3.47$). However, there were no significant interactions between the stimulus types and spatial frequencies, $F(2, 30) = 0.49$, $p = 0.599$, $\eta p2 = 0.032$, or among the task demands, stimulus types and spatial frequencies, $F(2, 30) = 1.21$, $p = 0.302$, $\eta p2 = 0.075$.

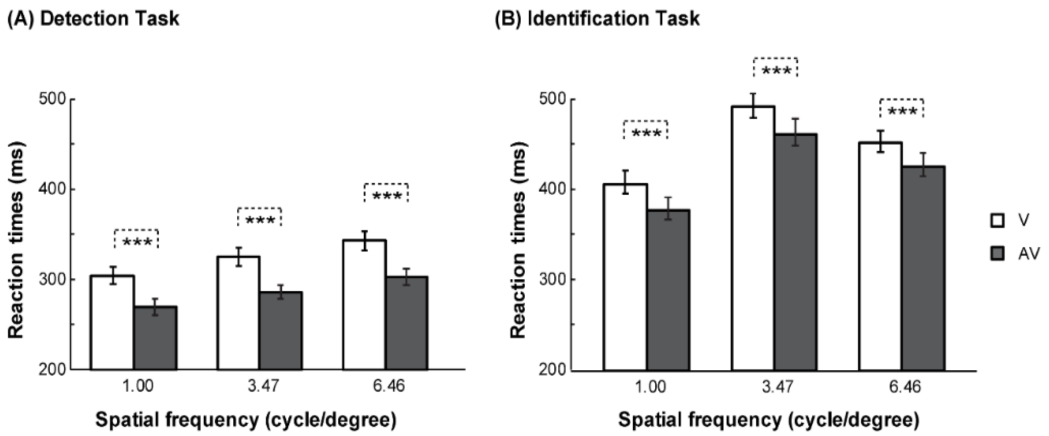


*Fig 3.2 Mean RTs during the detection task (A) and the identification task (B), *** p <0.001.*

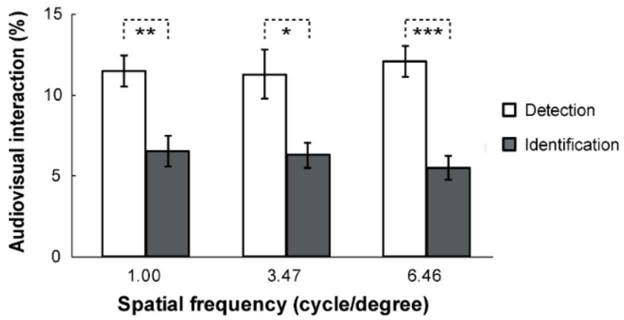


*Fig 3.3 Mean audiovisual interaction under three spatial frequencies in detection task and identification task, * p <0.05, ** p <0.01, *** p <0.001.*

Regarding the auditory facilitation benefits for each task, the data were reexamined in terms of audiovisual interactions (i.e., the difference between visual and audiovisual in each kind of spatial frequency; see Figure 3.3). The 2 task demands × 3 spatial frequencies ANOVA revealed a significant main effect of the task demands, $F(1, 15) = 31.19$, $p < 0.001$, $\eta p2 = 0.675$, with a stronger audiovisual interaction during the detection task than during the identification task. Further pairwise comparisons found that the audiovisual interaction during the visual identification task (6.11%) was weaker than that during the visual detection task (11.61%) for all spatial frequencies (all $p < 0.01$). However, no significant main effect for spatial frequencies ($F(2, 30) = 0.04$, $p = 0.939$, $\eta p2 = 0.003$), or interaction between the task demands and spatial frequencies ($F(2, 30) = 0.47$, $p = 0.533$, $\eta p2 = 0.03$) were found.

3.4 Discussion

In this study, we examined the audiovisual interaction elicited by visual detection and identification, and the results showed that significant audiovisual interaction occurred in both the detection and discrimination tasks (see Figure 3.2). However, the audiovisual interaction was mediated by task demands but not by spatial frequencies (see Figure 3.3).

The task-irrelevant auditory stimuli enhanced visual perception, regardless of the task demands or spatial frequencies (see Figure 3.2). The tasks replicated a classical cross-modal audiovisual interaction that has been reported previously in behavioural [58, 69], fMRI [70], ERP [71] and TMS [72] studies, suggesting that the simultaneous presence of visual stimuli and task-irrelevant auditory stimuli lead to enhanced processing of visual stimuli. Neurophysiological studies have reported that the auditory input activates the primary auditory cortex within 15 ms of presentation and then transmits the information to the visual cortex [73]. Falchier et al. (2002) systematically investigated multisensory processing in the primate striate cortex using anatomical methods [73]. Their results showed that the auditory input connected to the visual peripheral area directly and enhanced the visual cortex excitability. Subsequently, this connectivity was also found in the central visual field of humans [74]. Therefore, the low-level cortical interaction might lead to enhancements in visual perception. In addition, the integration between auditory and visual has also been found to occur via higher-order association cortices, such as the prefrontal and parietal cortices [70, 75]. Therefore, it is reasonable that task-irrelevant auditory stimuli facilitate visual perception.

The audiovisual interaction during the visual detection task was stronger than that during the visual identification task (see Figure 3.3). This result may be related to attention, which has been found to play an important role in audiovisual interaction. Mozolic et al. (2008) studied audiovisual integration using selective attention (attention to a single modality, visual or auditory) and divided attention (paying attention both visual and auditory), and their results indicated weaker audiovisual interaction with selective attention than with divided attention [76]. Therefore, it is possible that the attenuated audiovisual interaction in the present study was due to the further selectivity of the spatial frequencies. As is well known, the amount of information that can be attended to at once is limited the visual modality [77]. Investigations in visual studies using behavioral and ERP methods showed that visual identification utilizes additional attentional processes beyond those required for detection [66, 78]. Thus, it seems that some attention, which is necessary for audiovisual processing, shifted to identify whether the visual stimulus was the target spatial frequency, resulting in a reduced audiovisual interaction. Recently, Gibney et al. (2017) provided further evidence for the impact of attention on audiovisual interaction via studies using quickened audiovisual detection task and McGurk tasks using dual-task paradigms [79]. They also found decreased audiovisual integration and a reduced McGurk effect due to the decreased attention (diverting attention to the secondary task). Therefore, the attenuated audiovisual interaction during the identification task in this study could mainly be attributed to the decline in attention, which is needed for audiovisual processing.

The audiovisual interaction was not mediated by spatial frequency in either the detection task or the identification task (see Figure 3.3). This result extended Perez-Bellido's findings that a high contrast (83.3%) leads to the same amount of auditory enhancement of visual detection, regardless of spatial frequency [80]. This occurrence is possibly associated with early audiovisual integration, which is considered an automatic tendency to improve behavioural perception. Previous studies of humans demonstrated early audiovisual interaction in detection [23] and identification task [63]. Senkowski et al. (2011) reported the existence of early audiovisual integration and further clarified that early audiovisual integration was stimulus-driven processing, with a low-intensity stimulus resulting in stronger early audiovisual interaction [71]. In the present study, the extent of early audiovisual interaction between the visual and auditory stimuli was influenced by spatial frequency, but the use of a high contrast resulted in equivalent audiovisual interaction in both the detection (11.5%, 11.3%, 12.1% for 1.00, 3.47 and 6.46, respectively) and identification task (6.5%, 6.3%, 6.5% for 1.00, 3.47 and 6.46, respectively). Studies by Romei et al. (2013) provided further evidence by using TMS methods to test the time course of the cross-modal impact of looming sounds on visual perception, and they found that the attentional preferences of the participants affected the late stage, but not the early stage, of excitability changes [72]. Furthermore, De Meo et al. (2015) further proposed that early audiovisual interaction are a hallmark of bottom-up audiovisual processes that facilitate perception and behaviour directly, independent of task control [81]. Therefore, it is reasonable that the spatial frequencies did not modulate audiovisual interaction in either the detection or identification task.

3.5 Conclusions

In summary, our results demonstrated the importance of task demands and spatial frequencies for cross modal audiovisual interaction. Overall, visual detection rather than visual identification, was more beneficial for optimizing audiovisual interaction. These findings suggested that detection and identification modulated audiovisual interaction by relying on partially different mechanisms. However, we can’t dissociate the audiovisual interaction between detection and identification due to the lack of modulation by spatial frequencies, further electrophysiological studies are needed to confirm it.

Chapter 4 Visual intensity-dependent modulation: Effect of spatial frequency on audiovisual interaction in visual detection task

Summary

Although previous studies have shown that the auditory facilitation of visual detection is influenced by stimulus features, the impact of visual spatial frequency on the auditory facilitation of visual detection is still unclear. To examine the influence of spatial frequency on the auditory facilitation of visual detection, we designed a visual detection task with a task-irrelevant auditory stimulus while varying spatial frequency. The results showed that spatial frequency modulates the auditory facilitation of visual detection at low contrast (20%), but not at high contrast (100%). Moreover, the data revealed that the auditory facilitation of visual detection was larger for low and high spatial frequencies, and smallest at a spatial frequency of 0.70 cycles/degree. However, when visual contrast was adjusted to the same visual intensity, no significant difference was found among spatial frequencies. There was a significant interaction between spatial frequency and contrast during the auditory facilitation of visual detection, showing that lower visual intensity lead to greater auditory facilitation effect. These findings suggest the modulation of spatial frequency on the auditory facilitation of visual detection was dependent on visual intensity.

4.1 Background

In daily life, we perceive the environment through multiple sensory modalities, such as visual, auditory, tactile and so on. For example, when watching a movie, it is not only a visual experience, but auditory as well; the combination of visual and auditory information makes the movie more interesting and easier to understand. This ability of sensory integration or interaction is an essential component for detection. Many studies have shown that visual information detection is enhanced by auditory input, regardless of whether the input is relevant or irrelevant [69, 82, 83]. This facilitation effect is called "auditory facilitation of visual detection".

Indeed, the auditory facilitation of visual detection strongly depends on stimulus features, such as the frequency and intensity of the auditory stimulus [84, 85]. Visual stimuli have two basic features: contrast and spatial frequency. For contrast, behavioral studies have demonstrated that lower visual contrast leads to greater auditory facilitation of visual detection than higher visual contrast [80, 86, 87]. Moreover, the neural processing mechanism has been researched using event-related potential (ERP) and functional magnetic resonance imaging (fMRI) methods, with results suggesting stronger brain activity in response to lower contrast stimuli but not higher contrast stimuli [23, 71, 87], which is consistent with the inverse effect of multisensory integration (so-called *inverse effectiveness rules*). For spatial frequency, some behavioral studies have demonstrated the modulation of spatial frequency on audio-visual simultaneous judgment [88], fission illusion [89] as well as visual searching with a matched auditory stimulus [90]. However, relatively little is known about the influence of spatial frequency on the auditory facilitation of visual detection. One behavioral study reported that a simultaneous auditory stimulus can facilitate low spatial frequency detection in low (but not high) contrast [91]. However, they considered contrast and spatial frequency together and thus did not directly separate the effect of spatial frequency on the auditory facilitation of visual detection. Although much is known about the influence of visual contrast on the auditory facilitation of visual detection, whether and how spatial frequency modulates the auditory facilitation of visual detection remains unclear.

Here we performed three experiments using different contrast conditions to determine whether and how spatial frequency modulates the auditory facilitation of visual detection. In Experiment 1, we used a high contrast condition (100%) to test the effect of spatial frequency on the auditory facilitation effect. In Experiment 2, we tested the influence of spatial frequency on the auditory facilitation effect in a low contrast (20%) condition. In the Experiment 3, we tested whether visual detection was dependent on visual intensity, and visual intensity was adjusted by contrast to obtain an 80% threshold. To explore the auditory facilitation of visual detection among different spatial frequencies, reaction times, and hit rate were computed for visual stimulus and audiovisual stimulus over all spatial frequencies. Our results suggest that the modulation of spatial frequency on the auditory facilitation of visual detection depends on visual intensity.

4.2 The effect of spatial frequency on audiovisual interaction in a high contrast condition

The aim of this experiment (Experiment 1) was to investigate whether visual spatial frequency influences audiovisual interaction using a task-irrelevant auditory stimulus in a high-contrast (100%) condition.

4.2.1 Methods

4.2.1.1 Participants

Eighteen healthy volunteers (age range, 21–38 years; mean, 25 years old) participated in this experiment. All the participants had normal or corrected-to-normal vision and audition and were right-handed. The participants provided written informed consent for their participation in this study, which was previously approved by the ethics committee of Okayama University.

4.2.1.2 Stimuli

The visual (V) stimulus was a Gabor patch with vertical gratings (2-degree sinusoidal gratings enveloped by a Gaussian function; stimulus contrast was 100%), and the spatial frequency of the gratings included 0.54, 1.00, 1.86, 3.47 and 6.46 cycles/degree. The Gabor patch was presented in sine phase and was corrected for the monitor gamma. The visual stimulus was presented for 40 ms approximately 4° below the fixation point. The auditory (A) stimulus was a high-frequency (3 kHz, 65 dB) pure tone, which has been suggested to be processed early or integrated when accompanied by a visual stimulus [84]. The auditory stimulus was presented for 40 ms through a pair of headphones with a linear rise and fall time of 5 ms. The audiovisual (AV) stimulus consisted of a visual stimulus and an auditory stimulus, in which the auditory stimulus was accompanied by a visual stimulus of varying spatial frequencies, and the audiovisual stimulus was also presented for 40 ms.

4.2.1.3 Procedure

The experiment started with 5 minutes of dark adaptation. At the beginning of each session, the subjects were presented with a fixation cross for 3,000 ms. Following the fixation, there was an inter-stimulus interval (ISI) that randomly varied from 1,200 to

1,800 ms for subject response and rest. After the ISI, the visual stimulus, auditory stimulus and audiovisual stimulus were presented for 40 ms randomly, and the subjects were instructed to detect whether a visual stimulus was presented by pressing the left mouse button (Figure 4.1). Each subject completed five sessions in this study, with each session lasting approximately 6 minutes. Each session consisted of 150 trials in total, with 50 trials for each stimulus type.

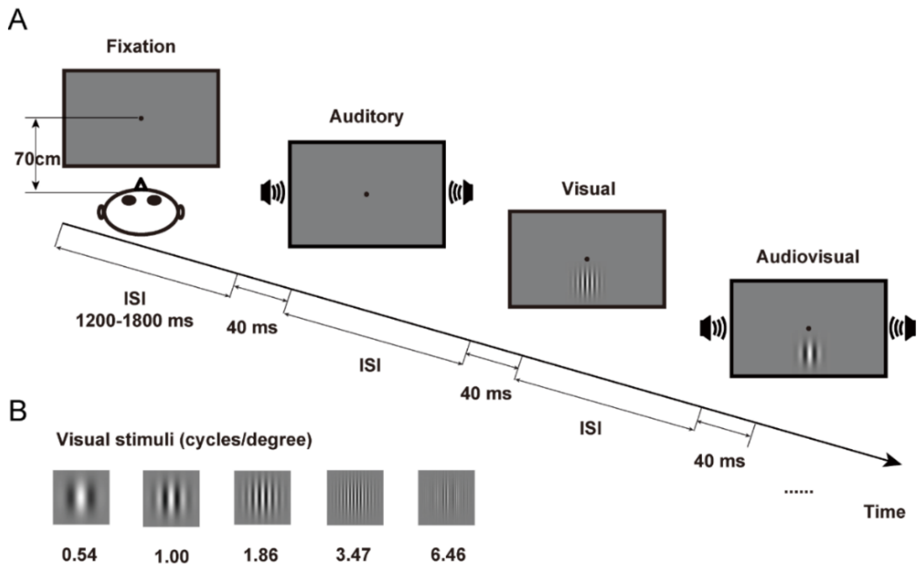


Figure 4.1 Schematic description of Experiment 1. (A) An example of the trial sequence in Experiment 1. After a fixation for 3000 ms at the beginning of each session, auditory stimuli, visual stimuli, and audiovisual stimuli were presented randomly with a random inter-stimulus interval of 1200 to 1800 ms. After a presentation of 40 ms for each stimulus, the subject needed to identify whether he/she observed a visual stimulus by pressing the left button. (B) The visual stimuli for the five spatial frequencies, namely, 0.54, 1.00, 1.86, 3.47, and 6.46 cycles/degree.

4.2.1.4 Apparatus

The experimental stimuli were generated and controlled by a custom-made program in MATLAB using the Psychophysics Toolbox [67] and a PC (XPS720, Dell; OS: Windows 10, Microsoft). The visual stimuli were displayed on a 100 Hz 17-in CRT monitor with a resolution of 1,280×1,024 pixels and a background luminance of 10 cd/m^2. Using a display attenuator that combines two 8-bit output channels of the graphics cards, the display system produced a 12-bit grey-level resolution (Cambridge Research Systems, Kyodo University) and was gamma corrected. The auditory stimuli were conveyed through headphones (MDR-1RNC, Sony, Japan). The participants viewed the monitor binocularly at a distance of 70 cm, with their heads stabilized on a chin rest.

4.2.1.5 Data analysis

Hit rates and RTs were computed separately for each subject and for each kind of stimulus. The hit rate was defined as the number of correct responses to the stimuli divided by the total number of stimuli. RTs were measured based on the timing of the participant's response to the presented stimulus. RTs that differed more than 3 SDs from the mean for each participant in each condition were excluded from analysis. Differences in the RTs and hit rates of the participants were analysed using a 2 stimulus type (V, AV) × 5 spatial frequency (0.54, 1.00, 1.86, 3.47, and 6.46) repeated measures analysis of variance (ANOVA). The Greenhouse-Geisser epsilon correction was used for non-sphericity, and the level of significance was fixed at a corrected $p < 0.05$. In addition, we further calculated the amount of improvement in the RTs (Sommers et al. 2005; Sumby and Pollack, 1954), and the auditory facilitation effect was expressed by formula (1).

$$\text{Auditory facilitation effect} = (\mathrm{RT}_{V} - \mathrm{RT}_{AV})/\mathrm{RT}_{V} \times 100 \qquad (4.1)$$

RT_{AV} is the RT of the correctly detected AV stimulus and RT_{V} is the RT of the correctly detected V stimulus. Differences in the auditory facilitation effect between the participants were analysed using one-way ANOVA with 5 spatial frequencies (0.54, 1.00, 1.86, 3.47, and 6.46). The Bonferroni correction was applied to the pairwise comparisons, and the level of significance was fixed at a corrected $p < 0.05$.

4.2.2 Results

Hit Rate: The mean hit rates for the V and AV stimuli are shown in Table 1 and Figure 4.2A. A 2 stimulus type (V, AV) × 5 spatial frequency (0.54, 1.00, 1.86, 3.47, and 6.46) repeated measures ANOVA on hit rates revealed a significant main effect of stimulus type (F (1, 17) = 5.20, p = 0.036, η_p^2 = 0.23). A pairwise comparison analysis showed that responses to the AV stimuli were more accurate than those to the V stimuli for the spatial frequency 6.46 (93.0% vs. 95.9%, p = 0.019), and a similar trend was observed for the spatial frequency 0.54 (97.8% vs. 98.8%, p = 0.083). The main effect of spatial frequency was also significant (F (4, 68) = 7.58, p < 0.001, η_p^2 = 0.308). Further pairwise comparison analysis results revealed that the hit rate for the spatial frequency 6.46 was lower than that for the spatial frequencies 0.54 (p <0.05), 1.00 (p <0.05) and 1.86 (p <0.05). However, there was no significant interaction between spatial frequency and stimulus type (F (4, 68) = 1.51, p = 0.228, η_p^2 = 0.081).

Reaction Times: Mean RTs for V and AV are presented in Figure 4.2B. The mean RTs was entered into a 2 stimulus type (V, AV) × 5 spatial frequency (0.54, 1.00, 1.86, 347, 6.46) repeated measures ANOVA. The main effect of stimulus type was significant, F (1, 17) = 214.34, p < 0.001, η_p^2 = 0.927, showing that RTs for the AV stimulus were faster than for V stimulus. It also revealed a significant main effect of visual spatial frequency, F (4, 68) = 83.42, p < 0.001, η_p^2 = 0.831, indicating that the RTs increased with spatial frequency for both V or AV stimuli (spatial frequency for 0.54, 1.00, 1.86, 3.47, 6.46 in V was 309 ms, 311 ms, 317 ms, 333 ms, 349 ms; for AV was 275 ms, 276 ms, 278 ms, 292 ms, 308 ms, respectively), see Figure 4.2B. However, there was no significant interaction between spatial frequency and stimulus type, F (4, 68) = 1.84, p = 0.175, η_p^2 = 0.098.

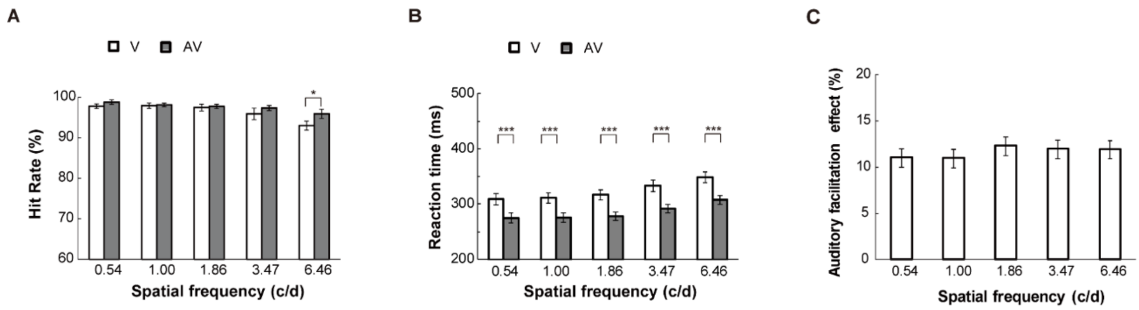


*Figure 3.2 Results of Experiment 1. (A) Mean hit rates for visual and audiovisual stimuli. (B) Mean RTs for visual and audiovisual stimuli. (C) Auditory facilitation effect. * p <0.05, *** p <0.001.*

Auditory facilitation effect: The mean auditory facilitation effect for each spatial frequency is presented in Figure 4.2C, with 11.00%, 10.94%, 12.28%, 11.94% and 11.89% for 0.54, 1.00, 1.86, 3.47 and 6.46, respectively. A repeated-measures ANOVA for the auditory facilitation effect revealed no significant main effect of spatial frequency, F (4, 68) = 0.602, p = 0.615, η_p^2 = 0.034.

4.2.3. Discussion

The results of this experiment show that, overall, a task-irrelevant auditory stimulus facilitated visual detection. The results indicate that responses to AV stimuli were faster than those to V stimuli for all spatial frequencies, and these results are in agreement with previous findings [61, 86, 92, 93] (Figure 4.2A). However, spatial frequency did not modulate the auditory facilitation of visual detection in the high contrast condition

(100%) (Figure 4.2C).

The auditory facilitation of visual detection was not influenced by spatial frequency: our results indicated that the magnitude of the auditory facilitation effect was equal in all spatial frequency conditions (11.0%, 10.9%, 12.3%, 11.9%, and 11.9% for 0.54, 1.00, 1.86, 3.47, and 6.46, respectively). This result is consistent with the findings from the study of Perez-Bellido (2013) [91]. Perez-Bellido et al. (2013) conducted a visual detection task by varying contrast for three spatial frequencies, in which they found an equivalent audiovisual enhancement (35 ms) for all spatial frequencies during a high contrast (82.2%) condition [91]. In the present study, the hit rate was greater for an audiovisual stimulus at a spatial frequency of 6.46, indicating that spatial frequency could alter audiovisual processing. However, the modulation of spatial frequency on the auditory facilitation effect did not occur; this might due to the high contrast, which resulted in a celling effect (the mean hit rates were 96% and 98% for the visual stimulus and audiovisual stimulus, respectively). Therefore, our results support the argument that spatial frequency does not modulate the auditory facilitation of visual detection during a high contrast condition [94, 95]. In the next experiment, we tested whether spatial frequency modulates the auditory facilitation of visual detection in a low contrast condition.

4.3 The effect of spatial frequency on audiovisual interaction in a low contrast condition

The aim of this experiment (Experiment 2) was to investigate whether spatial frequency influences audiovisual interaction by a task-irrelevant auditory stimulus in a low-contrast (20%) condition.

4.3.1 Methods

4. 3.1.1 Participants

Sixteen volunteers (age range, 21–31 years; mean, 25 years old) participated in this experiment. All the participants had normal or corrected-to-normal vision and audition and were right-handed. The participants provided written informed consent for their participation in this study, which was previously approved by the ethics committee of Okayama University.

4. 3.1.2 Stimuli

The experimental setup was the same as that of Experiment 1. However, in contrast to Experiment 1, a low visual contrast (20%) condition was used in the current experiment.

4. 3.1.3 Procedure

The procedure and task were the same as in Experiment 1.

4. 3.1.4 Data analysis

All data analyses were performed in an identical manner to those in Experiment 1.

4. 3.2 Results

Hit Rate: The mean hit rates for V and AV stimuli are presented in Figure 4.3A. A 2 stimulus type (V, AV) × 5 spatial frequency (0.54, 0.70, 1.00, 347, 6.46) repeated measures ANOVA was performed on the mean hit rate. The results revealed a significant main effect of stimulus type, F (1, 15) = 14.82, $p < 0.002$, $\eta_p^2 = 0.681$, showing a more accurate response when the visual stimulus was presented with a simultaneous auditory stimulus. The main effect of spatial frequency was also significant, F (4, 60) = 25.80, p < 0.001, η_p^2 = 0.653, thus indicating a less accurate response with increasing spatial frequency when spatial frequency was more than 0.70 cycles/degree. There was also a significant interaction between stimulus type and spatial frequency, F (4, 60) = 27.64, p < 0.001, η_p^2 = 0.174. The post hoc comparisons for stimulus type showed that the hit rates for the AV stimuli were higher than those for the V stimuli in the spatial frequencies 0.70 (*p* < 0.01), 3.47 (*p* < 0.05), and 6.46 (*p* < 0.01).

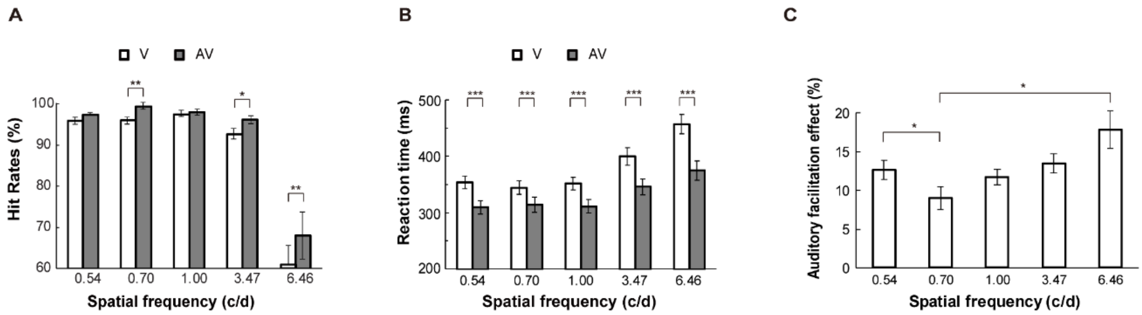


*Figure 4.3 Results of Experiment 2. (A) Mean hit rates for visual and audiovisual stimuli. (B) Mean RTs for visual and audiovisual stimuli. (C) Auditory facilitation effect. * p <0.05, *** p <0.001.*

Reaction Times: Mean RTs for V and AV are presented in Figure 4.3B. A 2 stimulus type (V, AV) × 5 spatial frequency (0.54, 0.70, 1.00, 347, 6.46) repeated measures ANOVA on the mean RTs revealed a significant main effect of stimulus type, F (1, 15) = 88.78, p < 0.001, η_p^2 = 0.885, indicating that responses to AV stimuli were significantly faster than those to V stimuli. The main effect of spatial frequency was also significant, F (4, 60) = 45.73, p < 0.001, η_p^2 = 0.680, indicating that the RTs were influenced by spatial frequency for both V and AV stimuli. Furthermore, the results also revealed a significant interaction between stimulus type and spatial frequency, F (4, 60)

= 73.15, $p < 0.001$, $\eta_p^2 = 0.830$, showing that the auditory facilitation of visual detection was different depending on the spatial frequency. The auditory facilitation effect is illustrated in Figure 3C. A repeated measures ANOVA on the auditory facilitation effect revealed a significant main effect of spatial frequency, F (4, 60) = 7.82, $p < 0.001$, $\eta_p^2 = 0.470$, revealing a weak auditory facilitation effect for the spatial frequency of 0.70. Additionally, with increasing (from 0.70 to 6.46) or decreasing (from 0.70 to 0.54) of spatial frequency, the facilitation effect became greater. These results suggest that changing spatial frequency influences the auditory facilitation of visual detection.

Auditory facilitation effect: The mean auditory facilitation effect for each spatial frequency is presented in Figure 4.3C, with 13.52%, 9.53%, 12.33%, 13.54% and 18.67% for 0.54, 0.70, 1.00, 3.47 and 6.46, respectively. One-way ANOVA revealed a significant main effect of spatial frequency on the auditory facilitation effect, F (4, 60) = 7.30, $p < 0.01$, $\eta_p^2 = 0.327$. A pairwise comparison analysis found that the auditory facilitation effect was smallest in the spatial frequency 0.70 condition and enlarged with the increase in spatial frequency (from 0.70 to 6.46) or with a decrease of spatial frequency (from 0.70 to 0.54) ($p < 0.05$).

4. 3.3. Discussion

The results indicated the presence of an auditory facilitation effect using a task-irrelevant auditory stimulus (Figure 4.3B). Furthermore, the auditory facilitation of visual detection was modulated by spatial frequency (Figure 4.3C).

The auditory facilitation of visual detection was influenced by spatial frequency in the low contrast condition (20%). This result is in agreement with the study by Perez-Bellido et al. (2013) in which a visual detection task was performed: they also found a modulation effect of spatial frequency on the auditory facilitation of visual detection

[91]. Their results showed that the strongest facilitation effect was only for the spatial frequency of 0.30 cycles/degree, but not for 5.93 cycles/degree. In addition, the modulation of spatial frequency on the auditory facilitation effect was reported in the 5% contrast condition but not for the 15.3% or 26.8% conditions. These differences might be caused by background luminance. In the study by Perez-Bellido et al. (2013), a relatively large background luminance was used (23.4 cd/m^2); we acknowledge that background luminance yields different results for spatial frequency detection [96]. Therefore, it is reasonable that spatial frequency modulated the auditory facilitation of visual detection.

However, the influence of spatial frequency on the auditory facilitation of visual detection was mediated by visual contrast. Experiment 2 found a spatial frequency modulation effect for the low contrast condition (20%), but not for the high contrast condition (100%) as in Experiment 1. This result is consistent with previous visual research that reported that visual contrast affected spatial frequency detection [95]. Huang et al. (2015) reported that visual processing was mainly based on visual intensity during a detection task [97]. Additionally, visual intensity-dependent modulation has also been found for audiovisual processing during visual detection [87]. Noesselt et al. (2010) reported that sound enhanced detection of low but not high intensity stimuli. In the present experiment, visual intensity was varied by spatial frequency, and the auditory facilitation effect was increased when spatial frequency was greater or less than 0.70 cycles/degree, which is in accordance with the inverse effectiveness principle of audiovisual integration. Therefore, it is possible that the modulation of spatial frequency on the auditory facilitation of visual detection was dependent on visual intensity.

4.4 The effect of spatial frequency on audiovisual interaction in same intensity condition

The purpose of this experiment (Experiment 3) was to investigate whether perceived intensity was the key factor for the auditory facilitation effect during visual detection. If the modulation is due to the perceived intensity, we expected an equivalent auditory facilitation effect when three spatial frequencies were adjusted to the same perceived intensity.

4. 4.1 Methods

4. 4.1.1 Participants

Eighteen volunteers (age range, 21–28 years; mean, 24 years old) participated in this experiment. All the participants had normal or corrected-to-normal vision and audition and were right-handed. The participants provided written informed consent for their participation in this study, which was previously approved by the ethics committee of Okayama University.

4. 4.1.2 Stimuli

The experimental setup was the same as that of Experiment 1. In Experiment 2, for the V stimuli, no significant difference was found among the spatial frequencies 0.54, 0.70 and 1.00 in terms of both hit rates and RTs. Therefore, 1.00, 3.47, and 6.46 were selected to conduct Experiment 3.

4. 4.1.3 Procedure

Prior to the experiment, all participants completed a preset designed to equate the detectability of the visual stimuli among different spatial frequencies. The preset paradigm was a single-interval go/no-go signal detection task. There were 6 sessions (2 sessions per spatial frequency), in which targets for each spatial frequency could appear at ten different contrast conditions including 2%, 2.5%, 3.96%, 4.45%, 5.6%, 6.28%, 7.79%, 10%, 15%, and 20%. Each session consisted of a total of 100 trials at 10 trials per condition. The participant was instructed to respond as they detected the visual stimulus by pressing the left mouse button. The 80% threshold was calculated using the 80% correct response rate for each subject in each spatial frequency. The three contrasts of the 80% threshold for each subject was used in the formal experiment, see Figure 4.4. The procedure and task of the formal experiment was same as Experiment 1, except a block design was used in this experiment.

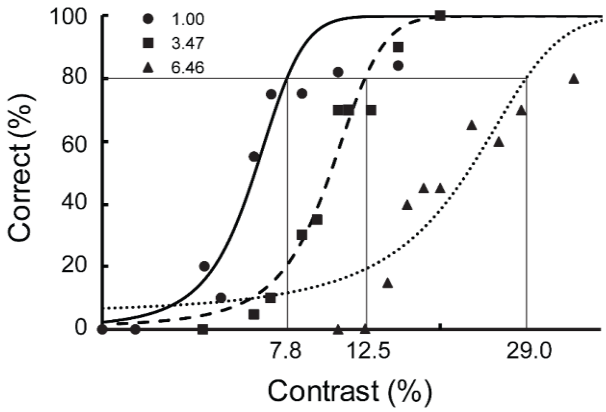


Figure 4.4 Detection accuracy as a function of contrast and spatial frequency; data taken from one participant. The circles, sequences, and triangle points represent the spatial frequencies of 1.00, 3.47, and 6.46, respectively. The 80% accuracy of the three spatial frequencies (7.8%, 12.5% and 29.0%) was used in the subsequent experiment.

4. 4.1.4 Data analysis

All the data analyses were performed in an identical manner to those of Experiment 1.

4. 4.2 Results

Hit Rate: Mean hit rates for V and AV stimuli are shown in Figure 4.5A. The performance in each case was close to the desired difficulty level of 80% correct for the V stimuli (1.00: 85.2%; 3.47: 81.6%; 6.46: 82.4%). A 2 stimulus type (V, AV) × 3 spatial frequency (1.00, 3.47, and 6.46) repeated measures ANOVA was performed on the mean hit rates. The results revealed a significant main effect of stimulus type ($F(1, 17) = 42.95$, $p < 0.001$, $\eta_p^2 = 0.716$). A pairwise comparison analysis indicated that the hit rates for the AV stimuli were significantly higher than those for the V stimuli at all spatial frequencies (all $p < 0.001$). However, no main effect of spatial frequency was found ($F(2, 34) = 0.74$, $p = 0.477$, $\eta_p^2 = 0.042$). Additionally, no significant interaction between stimulus type and spatial frequency was found ($F(2, 34) = 0.71$, $p = 0.497$, $\eta_p^2 = 0.040$).

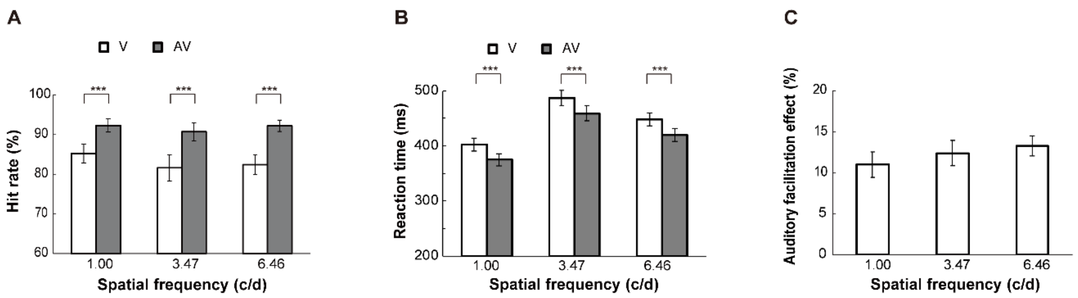


*Figure 4.5. Results of Experiment 3. (A) Mean hit rates for visual and audiovisual stimuli. (B) Mean RTs for visual and audiovisual stimuli. (C) Auditory facilitation effect. * p <0.05, *** p <0.001.*

Reaction Times: Mean RTs for V and AV stimuli are presented in Figure 4.5B. A 2 stimulus type (V, AV) × 3 spatial frequency (1.00, 3.47, 6.46) repeated measures ANOVA of the mean RTs revealed a significant main effect of stimulus type ($F(1, 17)$

= 91.49, $p < 0.001$, $\eta_p^2 = 0.843$). A further pairwise comparison found that the RTs to the AV stimuli were significantly faster than those to the V stimuli at all spatial frequencies (all $p < 0.001$). Furthermore, the results revealed a significant main effect of spatial frequency (F (2, 34) = 8.44, p = 0.002, η_p^2 = 0.332). A further pairwise comparisons found a faster response with 1.00 than with the other spatial frequencies ($p < 0.05$). Consistent with our hypothesis, there was no significant interaction between stimulus type and spatial frequency (F (2, 34) = 1.93, p = 0.161, η_p^2 = 0.102).

Auditory facilitation effect: The mean auditory facilitation effect for each spatial frequency is presented in Figure 4.5C, with 11.15%, 11.83% and 13.06% for 1.00, 3.47 and 6.46, respectively. One-way ANOVA revealed no significant main effect of spatial frequency on the auditory facilitation effect (F (2, 34) = 0.84, p = 0.435, ηp2 = 0.047).

4. 4.3. Discussion

The results indicated an auditory facilitation effect when using a task-irrelevant auditory stimulus. However, there was no significant difference in the auditory facilitation effect between the different spatial frequencies, see Figure 4.5 C.

Our results showed an equivalent auditory facilitation effect with three spatial frequencies (11.15%, 11.83% and 13.06% for 1.00, 3.47 and 6.46, respectively). This result was consistent with our hypothesis, showing that there was no significant difference in the auditory facilitation effect with varying spatial frequencies when the perceived intensity was equal. When the contrast was constant (Experiment 2), the perceived intensity was varied by spatial frequency. As a visual stimulus becomes weaker and the subject becomes more uncertain of his/her responses, the need for combining information from multiple modalities (i.e., visual and auditory) to form a super-additive multisensory response is also increased [98]. When the perceived intensity was adjusted to a constant level in this experiment, the need for combining auditory information was also equal, thus producing an equivalent auditory facilitation effect. Therefore, it is reasonable to suppose that the modulation of audiovisual interaction by spatial frequency was related to perceived intensity.

4.5 General discussion

In a series of experiments, we observed that responses to AV stimuli were faster and more accurate than responses to V stimuli, suggesting a significant auditory facilitation effect when accompanied by a task-irrelevant auditory stimulus. As indicated in neurophysiology studies conducted by Stein et al. (1993), the influence of a concurrent sound along with a visual stimulus evoked brain results from the integration of auditory and visual inputs by multisensory neurons in animals **[99]**. Recent behavior and neuroimaging studies have also suggested that audiovisual integration can enhance visual detection in humans. Jaekl et al. (2010) reported that a simultaneous sound improved visual contrast detection due to audiovisual integration **[100]**. The results of Li et al. further reported that a task-irrelevant sound improved visual discrimination resulting from audiovisual integration as examined using neuroimaging methods **[101, 102]**. These findings suggest that the mechanism underlying the auditory facilitation of visual detection results from integration between the auditory and visual modalities.

In this study, spatial frequency modulated the auditory facilitation of visual detection. We have described the auditory facilitation effect as a U-shaped function with the minimum occurring at a spatial frequency of 0.70 and is stronger for a low spatial frequency (0.54) and high spatial frequency (6.46). Such results are consistent with a study by Orchard-Mills et al. (2013), in which a U-sharped function was also found during visual search with a matched auditory stimulus [90]. These findings were correlated with visual contrast sensitivity [96, 103], which has been reported to have an inverted U-shaped function with peak contrast sensitivity occurring at a low spatial frequency (e.g., 2–3 cycles/degree). Neuroimaging studies found that BOLD signal strength increased with increasing perceptual sensitivity in the superior colliculus [104]; according to inverse effectiveness rules, a higher BOLD signal strength results in weaker audiovisual facilitation [87]. Therefore, our results provided the first evidence that spatial frequency influences the auditory facilitation of visual detection.

Additionally, the modulation of spatial frequency and contrast on the auditory facilitation of visual detection was dependent on visual intensity. The combined results of Experiments 1, 2 and 3 showed that the auditory facilitation effect decreased with increasing visual intensity. This observation fits with previous studies that demonstrated an inverse relationship between the magnitude of the auditory facilitation effect and visual intensity [85, 105]. Interestingly, this inverse association has also been recently found in animals, such as monkeys [106] and mice [107]. Neurophysiological studies have also reported this inverse effectiveness relevance to early multisensory processes [71]. To a certain extent, these findings were supported by the argument by Bizley et al. (2016), which stated that the integration between auditory and visual stimuli grouped stimuli features into auditory-visual objects during early audiovisual integration [108]. An animal study also provided evidence to clarify that this feature-binding effect occurred during audiovisual integration [109]. Therefore, it is reasonable that the influence of spatial frequency on the auditory facilitation of visual detection depends on visual intensity.

A number of studies have correlated early audiovisual integration with behavioral benefits [110, 111]. It therefore seems possible for early audiovisual integration in feature-binding (such as spatial frequency, size and contrast) to inform audiovisual decision making and result in improved visual detection. The findings reported here are in keeping with the argument of Daniel et al. (2014), which stated that no difference in auditory facilitation effect was observed when factoring out innate visual detection ability [112]. Thus, other audiovisual processing related with stimulus features may be interpreted by varying detectability, such as the different audiovisual behavior benefits for spatial location [69, 93, 113, 114], stimulus intensity [71, 85], and speed of visual motion [105]. However, in the present study, only present visual features were changed, including spatial frequency and/or contrast; thus, our study does not allow us to draw conclusions about stimuli feature-binding from auditory and visual influences of behavioral audiovisual facilitation. Further electrophysiological studies are needed to elucidate the neural mechanisms of integration under more detailed stimulus feature conditions.

4.6. Conclusions

In summary, we conducted three experiments to examine the modulation of spatial frequency on the auditory facilitation of visual detection. Our results showed that the influence of spatial frequency on the auditory facilitation of visual detection was mediated by contrast. When stimuli were adjusted to equal visual intensity for three spatial frequencies, the modulation of spatial frequency on the auditory facilitation of visual detection disappeared. Therefore, our study demonstrates that the modulation of spatial frequency on the auditory facilitation of visual detection depended on visual intensity.

Chapter 5 Effects of visual intensity on audiovisual interaction in discrimination task: an event-related potential study

Summary

A combination of signals across modalities can facilitate sensory perception. The audiovisual facilitative effect strongly depends on the features of the stimulus. Here, we investigated how spatial frequency, which is one of basic features of visual signal, modulates audiovisual integration with the event-related potential method (ERP). The behavioral results showed a significant audiovisual enhancement effect. Using event-related potential (ERP), audiovisual interaction in short latency ERPs with a left posterior topography were found for stimuli with 3.47 c/d, and the time window were delayed and smaller from auditory cortex (50-90 ms) to visual cortex (70-90 ms). These results extend findings from animal models to human visual cortices and highlight the impact of cross-sensory phase resetting by auditory stimulus on audiovisual interaction in ostensibly unisensory cortices. Moreover, audiovisual interaction was found over the frontocentral area for 1.00 c/d stimuli from 230–260 ms, for 1.86 c/d stimuli from 240–300 ms, for 3.47c/d stimuli from 280–320 ms. In addition, the audiovisual interaction also found in the parietal-occipital area for 1.00 c/d stimuli from 310–500 ms, for 1.86 c/d stimuli from 390–500 ms, for 3.47c/d stimuli from 480–500 ms. These findings suggest that a lower frequency visual signal paired with auditory stimuli maybe early processed or interaction despite the auditory stimuli being task-irrelevant information.

5.1 Background

In everyday life, our brain receives many sensory signals, such as vision or sound. The integration of information from different sensory modalities is an essential component for cognition. Previous studies have shown that responses to bimodal audiovisual stimuli are faster and more accurate compared with unimodal auditory or visual stimuli presented alone. This beneficial effect between visual and auditory stimuli is generally referred to as "audiovisual interaction".

Audiovisual interaction strongly depends on the intensity of auditory stimulus [115] and visual intensity [71]. It has been showed an inverse effectiveness rules, numerous studies have demonstrated that lower visual intensity leads to a greater audiovisual interaction than higher intensity stimulus [71, 87, 116]. Studies on spatial frequency-related audiovisual interactions mainly focus on simultaneous judgement, fission illusion, and visual searching rather than on visual detection/discrimination. As in our previous, auditory facilitated visual spatial frequency detection/discrimination depended on stimulus intensity (Study 2). Our results revealed that simultaneous presentation of an auditory stimulus and a visual stimulus can facilitate low spatial frequency detection in low contrast but that there was no significant facilitation effect in high contrast, and an equal auditory facilitation effect was obtained when the stimulus in same perceived intensity. Although much is known about the influence of visual intensity on audiovisual interaction with visual detection, the neural mechanism of stimulus intensity modulates audiovisual interaction remain unclear.

To investigate the effects of spatial frequency on audiovisual interaction in visual orientation discrimination task. Here, we studied the nature and timing of audiovisual interaction occurring with visual intensity in different spatial frequencies using the high temporal resolution of EEG. We found out fundamental patterns of influence of sound frequency, one of the basic characteristics of auditory stimuli, on audiovisual interaction.

5.2 Methods

5.2.1 Participants

Sixteen healthy volunteers (ages 22-29 years, mean age 24.1 years) from Okayama University participated in this experiment. All of the participants were right-handed, possessed normal or corrected to normal vision, and showed normal hearing ability. The experimental protocol was approved by Ethics Committee of Okayama University.

5.2.2 Stimuli

Visual stimuli were presented on a 17-in. CRT monitor (100 Hz, 1 280×1 024 pixels, the background luminance was 10 cd/m2) positioned 70 cm from the participant's head. The visual stimuli consisted of Gabor grating with three spatial frequencies:1.00, 1,.86, 3.47 cycle/degree (2° visual angle, 30% contrast), there were two kinds of which spatial frequency stimuli: clockwise 10° (standard/target) and anticlockwise10° (target/standard). These visual stimuli were presented approximately 4° below the fixation point and sustained 40 ms. The auditory stimulus consisted of a tone pip (40 ms in duration, 3000 Hz, 65dB SPL, 5 ms rise and fall periods), which presented through an earphone. This auditory stimulus was a task-irrelevant and ignored event. The audiovisual stimuli were consisted of simultaneous visual stimulus and auditory stimulus. There also two kinds of audiovisual stimuli: target and standard, which consistent with visual stimuli. The participants were required to discriminate the visual and audiovisual target stimuli.

5.2.3 Procedure and task

The study took place in a dimly lit, sound-attenuated, electrically shielded room. The participants sat on a comfortable chair and their head positions were fixed with a chin rest. Each participant participated in 12 blocks, of which each block lasted about 6 min, and 2 min rest between blocks. In the first six blocks, clockwise 10° was defined as target visual stimulus and anticlockwise10° was defined as standard, reverse in the other half of blocks, the order of orientation for target stimulus was balanced among participants. Each block consisted of 54 visual stimuli (45 standards, nine targets), 54 audiovisual stimuli (45 standards, nine targets), 15 auditory stimuli and 15 catch trials, and presented randomly. The inter-stimulus interval varied randomly between 800 and 1200 ms (mean 1000 ms). During the experiment, as shown in Figure 1, participants were instructed to fixate the fixation point and response to visual target stimuli as quickly and accurately as possible using their right hand, regardless of whether an auditory stimulus was presented. The participants were instructed to press the right button of a computer mouse when the orientation of target stimulus was clockwise and right button when the orientation of target stimulus was anticlockwise.

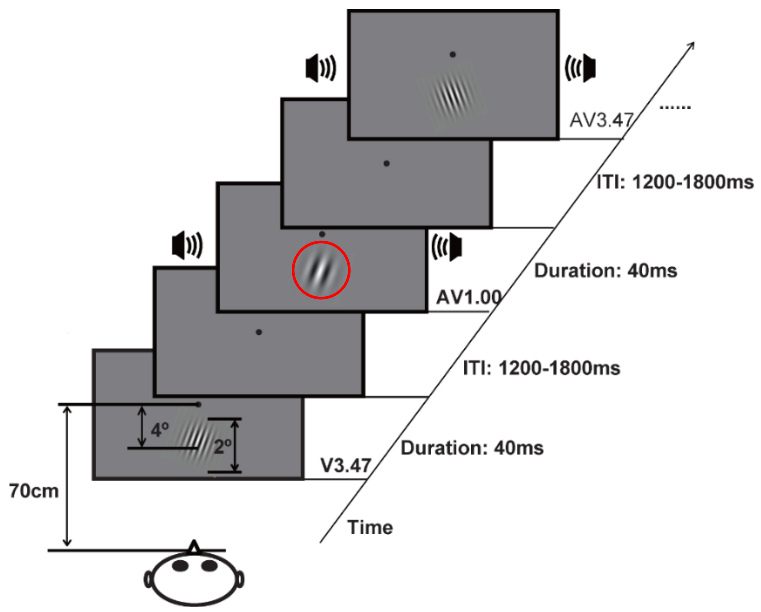


Figure 5.1. Experimental design. Stimuli were presented in a random stream of A, V and AV stimuli. The orientation of target stimulus was clockwise 10° in this example.

5.2.4 Apparatus

Stimulus presentation was control by running MATLAB 14th with Psychophysics Toolbox (Reference). An EEG system (BrainAmp MR plus, Gilching, Germany) was used to record EEG signals through 32 electrodes mounted on an electrode cap (Easy Cap, Herrsching-Breitbrunn, Germany), see Figure 5.2. All signals were referenced to left and right earlobe. Horizontal eye movements were measured by deriving the electrooculogram (EOG) from one electrode placed about 1cm from the outer canthi of the left eye. Vertical eye movements and eye blinks were detected by deriving an EOG from an electrode placed approximately 1.5 cm below the subject's left eye. The impedance was maintained below 5 kΩ. Raw signals were acquired at a sample rate of 500Hz and stored for off-line analysis.

5.2.5 Data analysis

5.2.5.1 Analysis of behavioral data

The hit rate was the percentage of correct responses relative to the total of target stimuli, and the false alarm (FA) were the percentage of the incorrect responses relative to the total of task-irrelevant stimuli. The mean reaction times (RTs) were calculated based on the correct responses that fell within the average time period ±2.5 SD. In addition, perceptual sensitivity (*d'*) and response criteria (c) were computed separately for different conditions (Stanislaw and Todorov, 1999). All of the results (Hit rate, FA, RTs, *d'*, c) were then analyzed using repeated-measures analysis of variance (ANOVA, Greenhouse-Geisser corrections with corrected degrees of freedom) at a significance level of 0.05, and the effect size estimates η_p^2 are reported.

5.2.5.2 Off-line analysis of EEG data

The ERPs elicited by the task-irrelative stimuli were analyzed by using the Brain Vision Analyzer software (version 1.05, Brain Products GmbH, Munich, Germany). The data were band-pass filtered from 0.01 to 60 Hz. Then, the data were divided into epochs, from -100 ms before stimulus onset to 600 ms after stimulus onset, and baseline corrections were made to the data from -100 ms to stimulus onset. Epochs contaminated by artifacts (i.e., eye movements, eye blinks, amplifier blocking, or false alarm) were rejected based on a threshold of ±80 µV in all channels before averaging. All averaged ERP waveforms were then digitally filtered of 0.1-30 Hz, and the grand-averaged data were obtained across all participants for each stimulus type in each electrode. Audiovisual interaction to the stimuli of 1.00, 1.86, and 3.47 was assessed by the difference wave [AV-(A+V)], which is obtained by subtracting sum of the ERP waves of the unisensory stimuli from the ERP waves of bimodal stimuli.

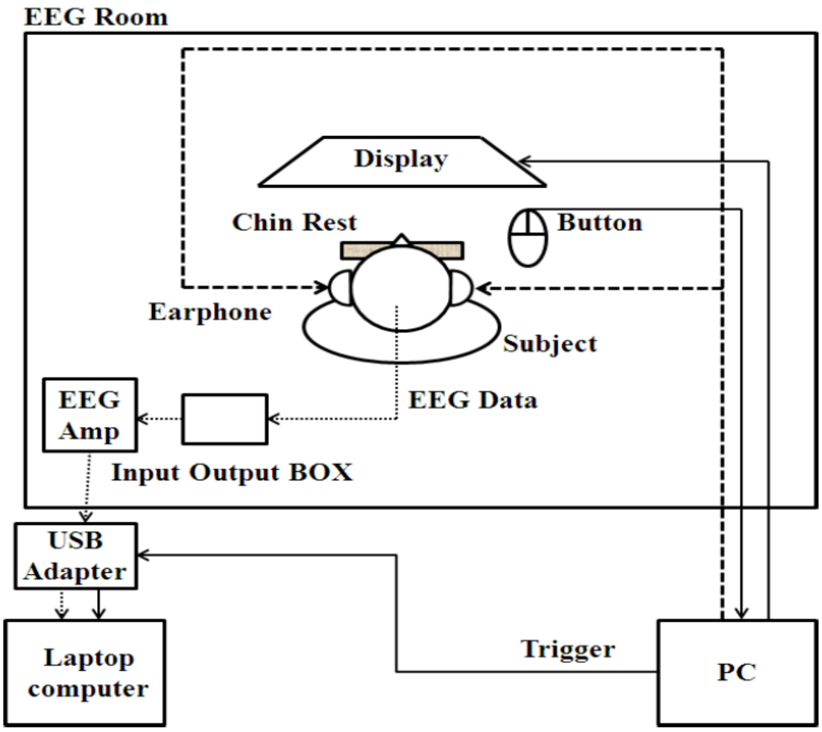


Figure 5.2 EEG systems

To establish the presence of audiovisual interaction, the statistical analysis was conducted in three steps. First, the ERPs for AV were compared with the linear summation of ERPs [A+V] using point-wise running t-test (two-tailed) for each electrode under each spatial frequency (1.00, 1.86, 3.47). A significant audiovisual interaction was defined as at least 10 consecutive data points met the alpha criterion of being < 0.05 (20 ms at 500 Hz digitization rate). Based on the results of t-test, we chose the three regions of interest (ROI) and time intervals when and where there was significant audiovisual interaction. Second, a 2 stimuli type (AV, A+V) × 3 spatial frequency (1.00, 1.86, 3.47) repeated-measures ANOVAs were performed for each ROI. If there were significant interaction among the stimuli type, spatial frequency, electrode and time interval found in the mean amplitudes, the third step of the analysis was conducted. In the third step, ANOVAs were conducted separately for each of the three spatial frequency in each time interval. The significance level was at 0.05, and the effect size estimates ηp2 were reported. The SPSS 16.0 software package (SPSS, Tokyo, Japan) was used for all statistical analysis.

5.3. Results

5.3.1 Behavioral results

RTs were shown in Table 5.1, A 2 stimulus type (V, AV) × 3 spatial frequency (1.00, 1.86, 3.47) repeated measures ANOVA on RTs showed that a main effect of stimulus type [F (1, 15) = 22.99, p = 0.000, η_p^2 = 0.605], indicating a faster response with AV stimuli than that with V stimuli. A main effect of spatial frequency [F (2, 30) = 17.33, p = 0.000, η_p^2 = 0.536] was also found. Further comparison found that RTs for 3.47 was slower than that for 1.00 (p = 0.001) or 1.86 (p = 0.000). In addition, there was also significant interaction between stimulus type and spatial frequency [F (2, 30) = 8.86, p = 0.001, η_p^2 = 0.371]. The post-hoc comparisons showed that the response with AV stimulus was significantly faster than that with V stimulus for spatial frequency 1.00 ($p < 0.01$) and 1.86 ($p < 0.001$), but not for 3.47 ($p = 0.867$). Moreover, repeated measures ANOVA on HA revealed significant main effect of stimulus type [F (1, 15) = 9.238, $p = 0.023$, $\eta_p^2 = 0.380$]. Further comparison found that the hit rate with AV stimulus was significantly larger than that with V stimulus for spatial frequency 3.47 ($p = 0.037$). In addition, repeated measures ANOVA on c also revealed a main effect of stimulus type [F (1, 15) = 5.31, $p = 0.036$, $\eta_p^2 = 0.262$], further comparison found a significant attenuated with AV stimulus in 1.86. The FA and *d'* were also shown in Table 5. 1. However, there was no significant main effect of stimulus type on FA [F (1, 15) = 2.24, $p = 0.155$, $\eta_p^2 = 0.130$] and *d'* [F (1, 15) = 1.18, $p = 0.295$, $\eta_p^2 = 0.073$] were found.

Table 5.1. Mean behavioral data for all participants in the experiment.

Stimulus types	RTs (ms)	HR (%)	FA (%)	d'	c
V1.00	533.6 ± 14.52	95.3 ± 1.14	0.71 ± 0.14	4.31 ± 0.15	0.36 ± 0.07
V1.86	542.9 ± 16.41	92.6 ± 1.86	1.04 ± 0.26	4.01 ± 0.17	0.39 ± 0.08
V3.47	581.6 ± 18.93	78.9 ± 4.61	0.67 ± 0.23	3.53 ± 0.18	0.76 ± 0.12
AV1.00	508.1 ± 15.77	95.6 ± 1.44	1.04 ± 0.22	4.29 ± 0.15	0.29 ± 0.08
AV1.86	497.1 ± 11.51	95.0 ± 1.21	1.47 ± 0.36	4.05 ± 0.14	0.26 ± 0.07
AV3.47	580.1 ± 17.73	84.8 ± 3.67	0.66 ± 0.22	3.73 ± 0.18	0.65 ± 0.09

5.3.2 ERP results

Evoked brain activity to unisensory stimuli: The ground-averaged ERPs to unisensory auditory stimuli and unisensory visual stimuli are shown in Fgure 5.2. The amplitude of the auditory evoked P1 (108 ms post-stimulus onset), N1 (182 ms), P2 (254 ms) components at Fz, see Fig 2A. For the three different spatial frequencies of visual stimuli (1.00, 1.86, 3.47), the ERPs showed a negativity-polarity wave peaking (N2) at around 256 ms (-0.44 µV), 270 ms (-0.85 µV), 298 ms (-1.33 µV) at Oz, respectively (Fig 2B). Apparently, the amplitude of the N2 component decreased with increasing spatial frequency, and the latency of the N2 delayed with increasing spatial frequency (1.00 vs. 1.86: $p < 0.05$; 1.00vs. 3.47: $p < 0.001$).

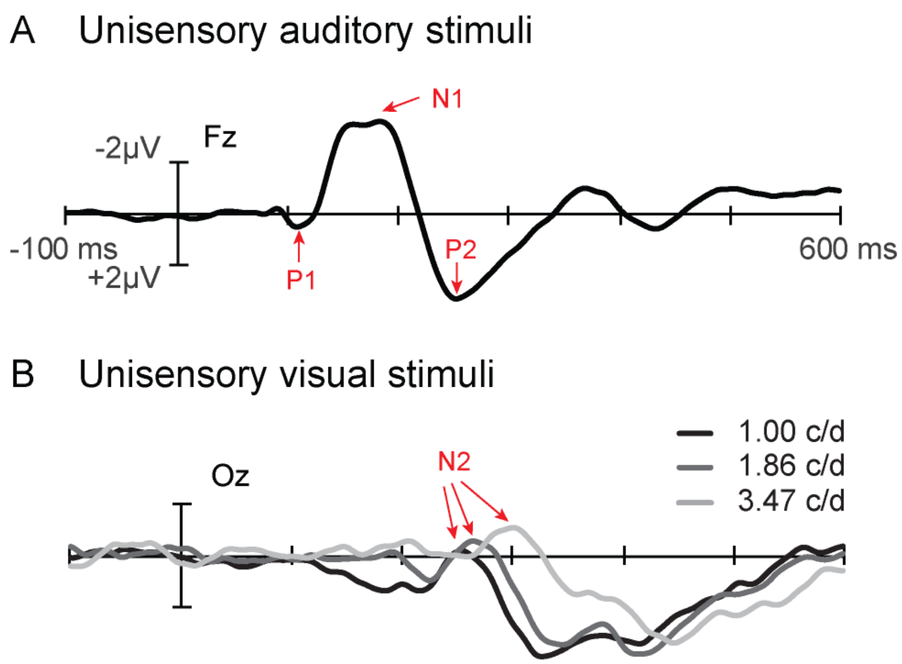


Figure 5.3. Waveforms of unisensory auditory (A) and visual stimuli (B).

Audiovisual interaction in the evoked brain activity: To examine the neural activities responsible for audiovisual interaction, the point-wise running t-tests (AV vs. A+V) for 1.00, 1.86 and 3.47 were computed. Based on the t-test statistical analysis and the topographical response pattern, three ROIs were selected. Left posterior area (F7, FC5, T7, C3, CP5, CP1, P7, P3 and O1) at 50-90 ms time interval. Fronto-central area (F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6) at 230-320 ms time interval. Parietal-occipital area (CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, Oz, O2) at 310-500 ms time interval. We analyzed these different audiovisual integration patterns for three spatial frequencies in detail as follows.

Early audiovisual interaction (50-90 ms) at left posterior area: Early audiovisual interactions have found for 3.47 condition in left posterior area at 50-90 ms time interval, see figure 5.5 and Table 5.2. In addition, it was clearly observed that the latencies of audiovisual interaction were delay and the time intervals of audiovisual interaction were smaller from electrode T7 (50-110 ms) to O1 (70-90 ms), see figure 5.4. Thus, we chose the 70-90 ms time interval for use in further analyses. Analysis of mean amplitudes using the 2 stimuli type (AV, A+V) × 3 spatial frequency (1.00, 1.86, 3.47) repeated-measures ANOVA revealed a trend interaction between stimuli type and spatial frequency [F (2, 24) = 3.94, $p = 0.052$, $\eta_p^2 = 0.247$]. Further post hoc analysis revealed significant audiovisual interaction effects for 3.47 ($p < 0.001$), and the amplitude was more positive in AV condition (0.39 µV) compared to [A+V] condition (-0.34 µV). However, the ANOVA for 1.00 and 1.86 did not reveal any significant effects. These results indicate that auditory enhanced high spatial frequency perception via direct connectivity from auditory cortex to visual cortex during early stage.

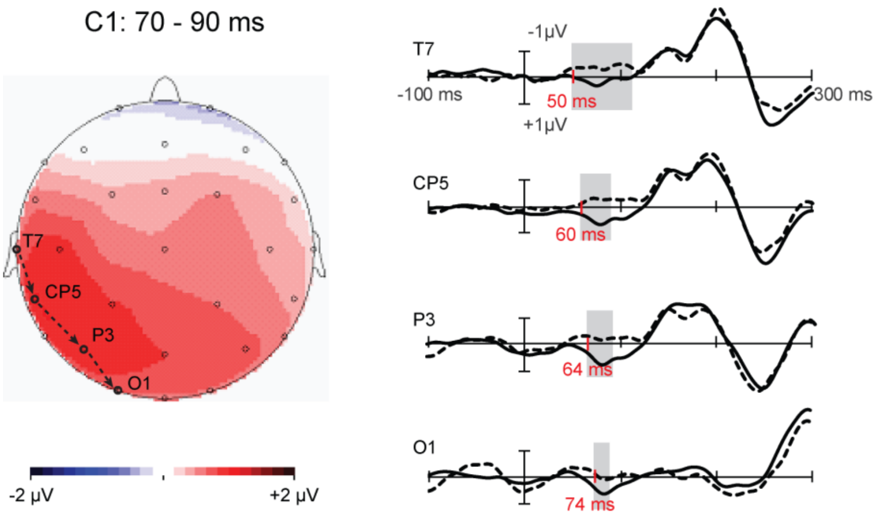


Figure 5.5. Topography of the different significant spatio-temporal patterns of interaction in the left posterior region. An obvious pattern of interaction effects was visible at 50-90 ms for 3.47 c/d.

Audiovisual interaction over frontocentral area at 240-320 ms: Audiovisual interactions at the time of onset and the latency of time interval were notably different in the three conditions at fronto-central area, see figure 5.6 and Table 2. Thus, ANOVAs were performed separately for three spatial frequencies using the factor of stimuli type. The results showed a significant main effect of stimuli type for spatial frequency of 1.00 in 230-260ms [F (1, 12) = 14.57, p = 0.002, η_p^2 = 0.548], 1.86 in 240-300ms [F (1, 12) = 13.26, p = 0.003, η_p^2 = 0.525], 3.47 in 280-320 [F (1, 12) = 9.54, p = 0.009, η_p^2 = 0.443], revealed a larger positivity for AV ERP than the summed [A+V] ERP waveform. These results showed that the time of audiovisual interaction onset were delayed with increasing spatial frequency (230 ms, 240 ms and 280 ms for 1.00, 1.86 and 3.47, respectively), suggesting a delayed audiovisual interaction effect with increasing spatial frequency.

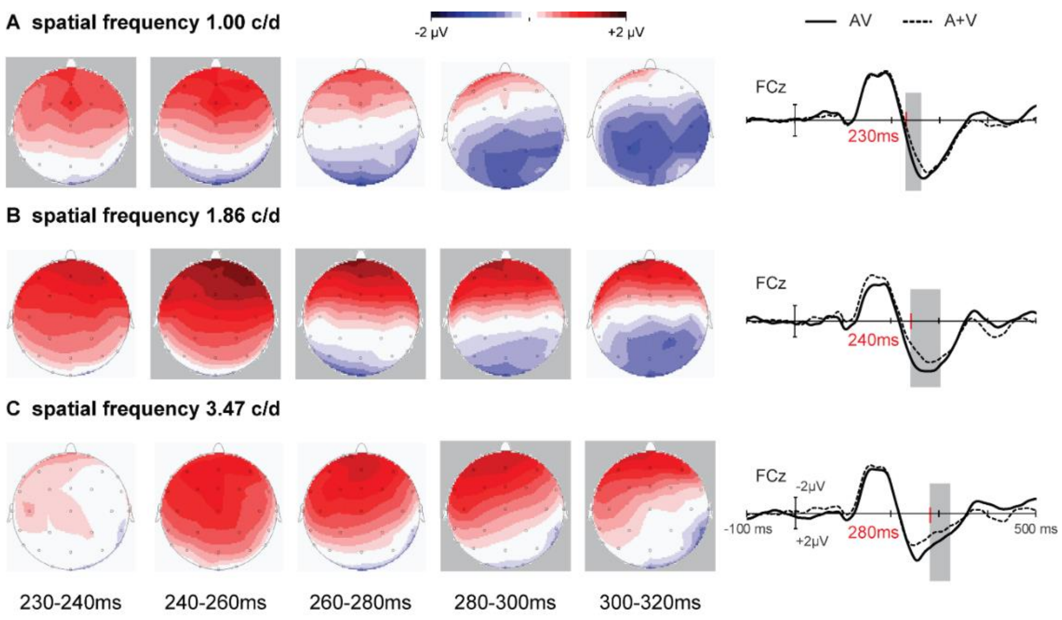


Figure 5.6. Topography of the different significant spatio-temporal patterns of interaction in fronto-central.

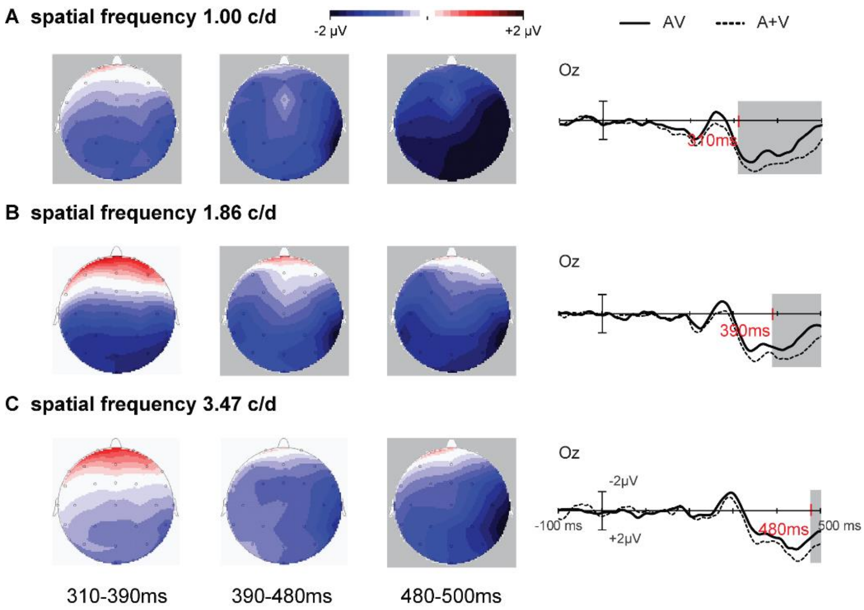


Figure 5.7. Topography of the different significant spatio-temporal patterns of interaction in parietal-occipital area.

Audiovisual interaction over parietal-occipital area at 310-500ms: Audiovisual interactions at the time of onset and the latency of time interval were notably different in the three conditions at parietal-occipital area, see figure 5.7 and Table 2. Thus, ANOVAs were performed separately for three spatial frequencies using the factor of stimuli type. The results showed a significant main effect of stimuli type for spatial frequency of 1.00 in 310-500 ms [F (1, 12) = 13.60, p = 0.003, η_p^2 = 0.531], 1.86 in 390-500 ms [F (1, 12) = 14.53, p = 0.002, η_p^2 = 0.548], 3.47 in 480-500 [F (1, 12) = 7.81, p = 0.016, η_p^2 = 0.394], revealed a larger positivity for the summed [A+V] ERP than AV ERP waveform. These results showed that the time of audiovisual interaction onset were delayed with increasing spatial frequency (340 ms, 390 ms and 480 ms for 1.00, 1.86 and 3.47, respectively), suggesting a delayed audiovisual interaction effect with increasing spatial frequency.

Table 5.2 the parameter of time interval, amplitude, and p-value in audiovisual interaction

Conditions	**Time interval (ms)**	**AV (M±SE)**	**[A+V] (M±SE)**	***p*-value**
Left posterior area				
1.00	70-90	-0.13 ± 0.16	-0.06 ± 0.22	0.833
1.86	70-90	-0.06 ± 0.17	-0.07 ± 0.12	0.953
3.47	70-90	0.39 ± 0.22	0.34 ± 0.18	0.001
Fronto-central area				
1.00	230-260	3.29 ± 0.25	2.18 ± 0.63	0.002
1.86	240-300	3.38 ± 0.24	2.58 ± 0.70	0.003
3.47	280-320	1.97 ± 0.63	1.13 ± 0.70	0.009
Pareital-occipital area				
1.00	310-500	3.11 ± 0.61	4.41 ± 0.94	0.003
1.86	390-500	2.85 ± 0.59	4.31 ± 0.90	0.002
3.47	480-500	2.19 ±0.55	3.43 ± 0.91	0.016

5.4. Discussion

We studied the effects of visual spatial frequency on audiovisual interaction in evoked brain activity with ERP method. The results of present study clearly showed that spatial frequency affects audiovisual interaction. An earliest audiovisual interaction (50–90 ms) with high spatial frequency was found over left posterior region. Besides, spatial frequency also modualted late audiovisual interaction. An obvious pattern of integration effect was visible at the fronto-central region and parietal-occipital region over approximately 230–320 ms and 310-500 ms, respectively, and occurred earlier when the spatial frequency was in low condition.

Audiovisual interactions in early evoked brain activity

Early audiovisual interaction has been observed (50-90 ms) and the onset of audiovisual interaction was delayed from electrode of T7 (50 ms) to electrode of O1 (74 ms), see figure 5.4. In line with previous findings of early audiovisual interactions around 50 ms [23, 117]. Investigation of anatomical study in monkey have reported the existence of indirect projection that auditory inputs reach areas of multisensory convergence (superior temporal gyrus/sulcus) and then transmitted via feedback connections to earlier visual areas [118]. Investigation of functional magnetic resonance imaging also have reported that multisensory interaction in human superior temporal sulcus and [116, 119, 120]. These studies suggested that posterior superior temporal sulcus (pSTS) plays an important role in audiovisual interaction. However, recently, numerous studies reporting the presence of direct feedforward projection from primary auditory cortices to early visual areas. Falchier et al. (2002) have proposed that auditory cortex connected with visual cortex directly in cat and the connectivity was greater in peripheral visual area than in the center visual area [73]. Similarly, direct projections from A1 to the visual cortex (area 21) have been shown in ferrets [121]. More recently, probabilistic fibre-tracking using diffusion MRI revealed fibre tracts between Heschl's gyrus and both anterior regions of the calcarine sulcus and also the occipital pole [122]. Therefore, we propose that the presence of connectivity from auditory inputs to the visual cortex caused the early audiovisual interaction and enhanced visual perception.

Moreover, audiovisual interaction occurred with high spatial frequency of 3.47 but absent in low spatial frequency of 1.00 and 1.86, see figure 5.4. This interaction phenomenon may be related to the stimulus intensity. Senkowski et al. (2011) reported an inverse relationship between early audiovisual interaction and stimulus intensity [71]. Their results found early audiovisual interaction only for low intensity inputs, but not for stimuli with middle and high intensity. Moreover, study of Talsma et al. (2007) also observed early audiovisual interaction by using related low intensity visual stimuli co-occurred with middle intensity auditory stimuli [123]. These studies suggested that early interactions occurred primarily at least one of the presented inputs was relatively low in intensity. In the present study, high spatial frequency was considered as low intensity stimulus, as an investigation of visual studies have proven that visual spatial frequency altered intensity when contrast was constancy [124]. Therefore, it is reasonable that early audiovisual interaction occurred with high spatial frequency.

In addition, lateralization effect occurred for early audiovisual interaction, see figure 5.4 and 5.5. Early audiovisual interaction only occurred in left posterior region but not right posterior region. This left lateralization effect have also been observed in the previous multisensory studies [71, 125, 126]. Calvert et al. (2000) have claimed that the region in left posterior superior temporal sulcus (pSTS) is more activated by audiovisual than unimodal processing [125]. Subsequently, Senowski et al. (2011) further reported an audiovisual interaction in left posterior region and interaction was followed inverse effectiveness [71]. Consistent with our results, their results showed that low intensity stimuli caused stronger audiovisual interaction in left posterior region. These results may relate with the ventral visual pathway which is associated with object recognition and form representation process. Investigation of visual studies have reported that high spatial frequency information projects chiefly to the ventral cortical visual stream [127]. In addition, previous fMRI studies have also reported that, when the baseline is fixation, activation is left lateralized in posterior ventral occipitotemporal for stimuli with high spatial frequency [128]. In the present study, participants were instructed to discrimination the orientation of visual grating, one feature of the visual stimulus, major process via ventral pathway. Therefore, it is reasonable that high spatial frequency resulting in early audiovisual interaction at left posterior region.

Audiovisual interaction in late evoked brain activity

Audiovisual interaction was significantly delayed with increasing spatial frequency in fronto-central region. Our results showed that audiovisual interaction occurred in the 230-260 ms time interval for 1.00 c/d, 240-300 ms time interval for 1.86 c/d and 280-

320 ms time interval for 3.47 c/d, see figure 5.7. The anatomical, physiological, transcranial magnetic stimulation (TMS) and neuroimaging studies have provided some support for the multisensory activity in prefrontal cortex (include dorsolateral prefrontal cortex and ventrolateral prefrontal cortex) [20, 129, 130]. Moreover, as in our results, Wu et al. (2015) observed audiovisual interaction occurred over fronto-central region during 300-340 ms time interval [84]. Therefore, it was acceptable that audiovisual interaction occurred at fronton-central region. Additionally, in the present study, the negativity-polarity wave peaking (N2) at around 256 ms (-0.44 μV), 270 ms (-0.85 μV), 298 ms (-1.33 μV) at Oz, respectively. The visual evoked brained activity was delayed with increasing spatial frequency, see Figure 5.3. The onset time of N2 peak consisted with the onset time of audiovisual interaction. Therefore, the delayed onset time of audiovisual integration might be mainly attributed to the slower processing speed of unimodal visual perception.

Furthermore, audiovisual interaction in parietal-occipital region was also delayed with increasing spatial frequency, see Figure 5.7. Audiovisual interaction occurred in the 310-500 ms, 390-500 ms and 480-500 ms tome interval for 1.00, 1.86 and 3.47 c/d, respectively. Previous work have pointed that audiovisual activations in parietal-occipital association regions by fMRI [131], and ERP [132]. Furthermore, Yang et al. (2015) reported that stimulus feature modulated the onset time of audiovisual interaction in occipital area [84]. Therefore, it is reasonable that spatial frequency modulated audiovisual interaction in parietal-occipital region. However, many studies have attributed the perceptual benefits to early sensory encoding in primary cortices. Gao et al. (2014) proposed that audiovisual interaction can occur at very early latencies (180-200 ms) relative stimulus onset over parietal-occipital areas [133]. Inconsistent with abovementioned findings, in the present study, the audiovisual interaction occurred at late stage of 300-500 ms time interval. Kayser et al. (2017) reported that sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations [134]. Therefore, the late visual orientation of spatial frequency representation might provide a reason for delayed audiovisual interaction.

5.5 Conclusions

The study confirmed that audiovisual interaction was greatly influenced by stimulus intensity, and low intensity stimulus produced an early audiovisual interaction over left posterior region. Additionally, the results showed a similar intensity-delayed effect on audiovisual interaction in fronton-central region and parietal-occipital region.

Chapter 6 General conclusion and future projections

Summary

This thesis has investigated the mechanism of cross-modal audiovisual interaction, and the diversity between visual detection and visual discrimination. Additionally, the visual intensity-related audiovisual interaction has also been evaluated. In this chapter, our findings are summarized below. Further, some future projections are included.

6.1 General conclusions

The current thesis includes four experiment studies. The first experiment is a leading of the thesis, investigating the effect of perceptual complex on visual processing and the diversity of response between detection and discrimination behaviorally. The second experiment to detect whether the perceptual complex affect audiovisual interaction and showed in which stage the diversity was presented. The third experiment examined the diversity of spatial frequency on audiovisual interaction in detail by visual detection task and showed in which stage the diversity was presented. Basing on the behavioral data, the stimulus intensity effect was found in audiovisual interaction. Therefore, in the fourth experiment, we designed an audiovisual interaction with visual discrimination task to evaluate the audiovisual interaction of stimulus intensity in detail by EEG.

Chapter 2 Describes the influence of perceptual process (detection and discrimination) on visual processing. A visual detection and visual orientation discrimination task were used to test the visual threshold with different spatial frequency. Our results showed that there was no significant difference in threshold between visual detection and visual discrimination, whereas the response time for visual detection were faster than that for visual orientation discrimination. These results suggested that the perceptual of detection and discrimination might rely on partially separate mechanisms.

Chapter 3 Describes the influence of perceptual process on audiovisual interaction, and the diversity of audiovisual interaction with varying spatial frequency. In this part, we designed a visual detection and visual discrimination task with/without a task-irrelevant auditory stimulus were conducted to examine the effect of perceptual process on audiovisual interaction, and the difference between different spatial frequency. The results confirmed that the response for visual discrimination was slowed and provided empirical evidence that visual discrimination attenuated audiovisual interaction. However, the audiovisual interaction was not mediated by spatial frequency in either the detection task or the discrimination task due to high contrast.

Chapter 4 Describes stage in which that the diversity of visual spatial frequency on audiovisual interaction. To clarify this, a visual detection task with/without a task-irrelevant auditory stimulus was performed. The results showed that spatial frequency modulates audiovisual interaction at low contrast (20%) but not at high contrast (100%) condition. Moreover, the data revealed that audiovisual interaction was larger for low (0.54 cycles/degree) and high (6.46 cycles/degree) spatial frequencies than for a medial spatial frequency of 0.70 cycles/degree (all $p < 0.05$). However, when the visual stimulus was adjusted to the same perceived intensity for each spatial frequency by changing contrast, no significant difference was found among the different spatial frequencies ($p > 0.05$). The current results suggested that the intensity of visual stimulus is the key factor for audiovisual interaction.

Chapter 5 Describes stage in which that the effect of stimulus intensity on audiovisual interaction was presented. To further clarify the effect of visual intensity on audiovisual interaction, the event-related potential (ERP) method was used. In this study, a visual orientation discrimination task with/without a task-irrelevant auditory stimulus was performed. The results showed that in the low intensity (spatial frequency of 3.47c/d) condition existing the earliest interaction (50 - 90 ms) in the left posterior region. This audiovisual interaction was delayed from auditory cortex (50-90 ms) to visual cortex (70-90 ms). This result indicated that auditory enhanced low intensity visual perception via directing connectivity from auditory cortex to visual cortex during early stage. Moreover, the audiovisual interaction over frontocentral area was found for all conditions and delayed with decreasing intensity (230-260 ms, 240-300 ms, 280-320 ms for three intensity). In addition, audiovisual interaction over parietal-occipital area were delayed with decreasing visual intensity (310-500 ms, 390-500 ms and 480-500 ms for the intensity of 1.00, 1.86 and 3.47 c/d). These results suggested that the audiovisual interaction pattern was different with visual intensity, and further revealed a delayed audiovisual interaction resulting from the slowed visual processing.

6.2 Future projections

Firstly, the current results of the thesis revealed that there is a significant diversity of audiovisual interaction with different stimulus intensity, including both the early stage and later stage of audiovisual interaction, and the brain regions. Therefore, one of the important challenges for the future studies is general development of audiovisual interaction across all the life span. Secondly, the cognitive processing altered greatly with aging, whether the alteration of audiovisual interaction with aging was due to the decline of cognitive processing is also important projection.

In addition, according to the current situation, future studies will focus on special populations (e.g. older people, patients with headache, mild cognitive impairment, Alzheimer's disease, and schizophrenia) to uncover the neural mechanism of audiovisual interaction and to provide important basis for the early clinical detection and rehabilitation of special brain disease.

Appendix

I Simple introduction of EEG and B++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, ***Figure AI-1***.

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 BrainAmp 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 AI-1. *EEG amplifier of BrainAmp MR plus*

Table AI-1. *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 / eference	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

The current thesis adapted 32 electrodes of this apparatus. The location and name of each channel that the present study was used is displayed in ***Figure AI*-2**.

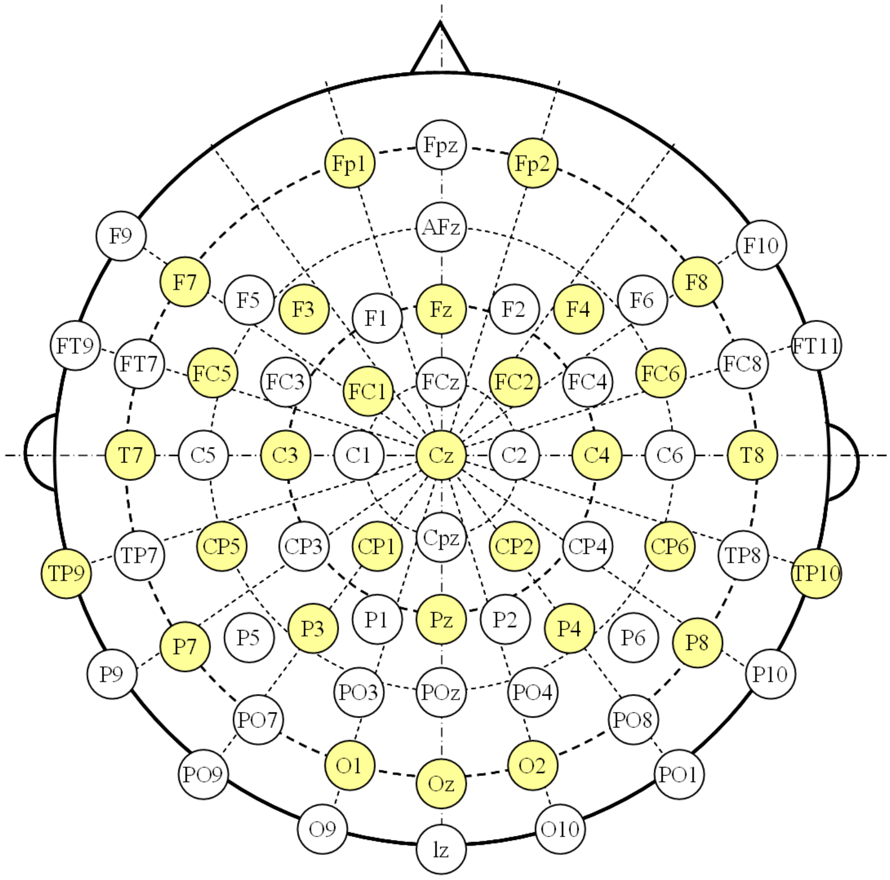
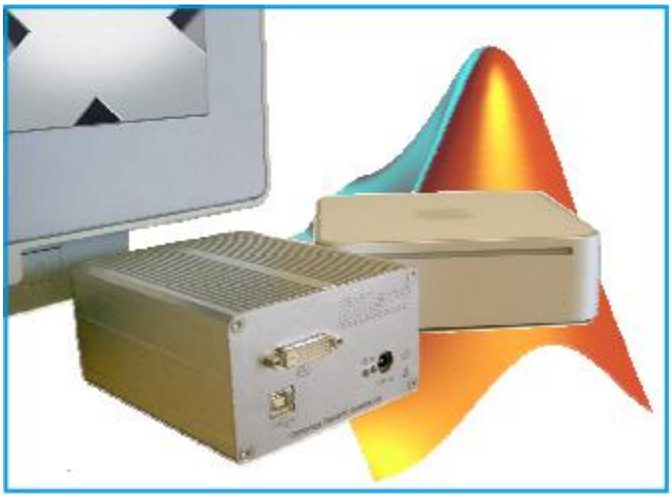


Figure AI-2. *The locations and names of each electrode.*

The current thesis adapted the resolution of the CRT screen, by a display attenuator that combines two 8-bit output channels of the graphics cards, the display system produced a 12-bit grey-level resolution (Cambridge Research Systems, Kyodo University)

(A)



(B)

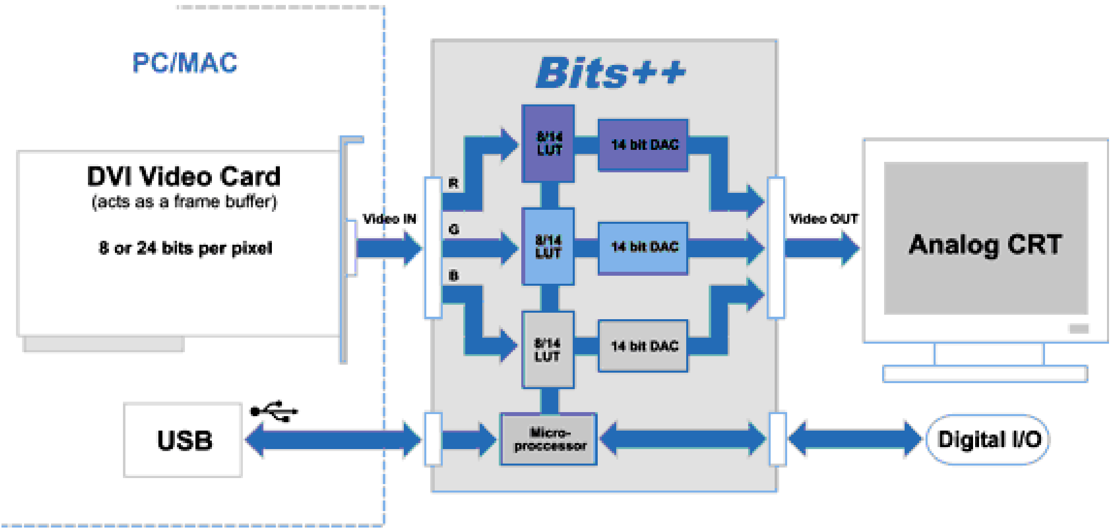


Figure AI-3. *B ++ for improve the resolution of CRT screen*

II ERP data in Experiment 4 (Chapter 5)

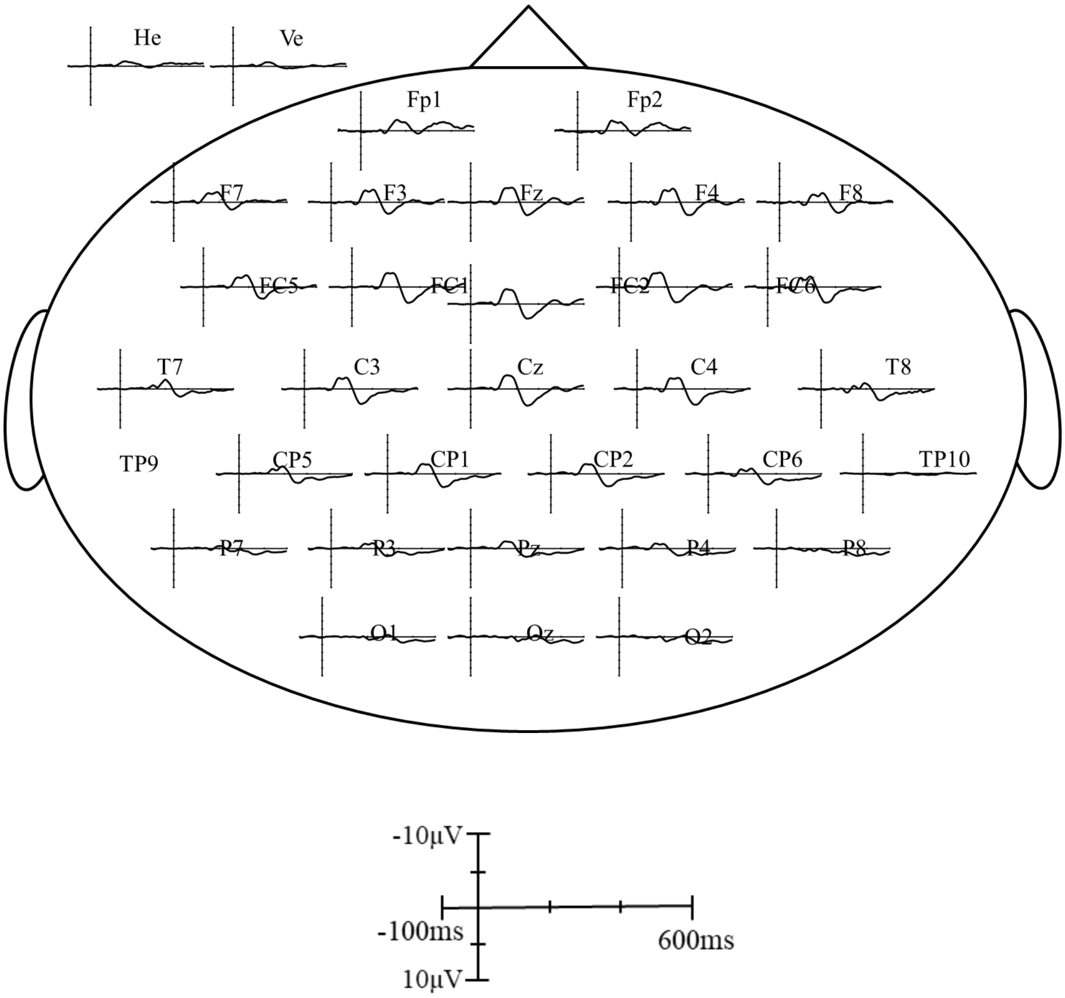


***Figure AII-1**. Averaged Event-Related Potentials of auditory stimulus*

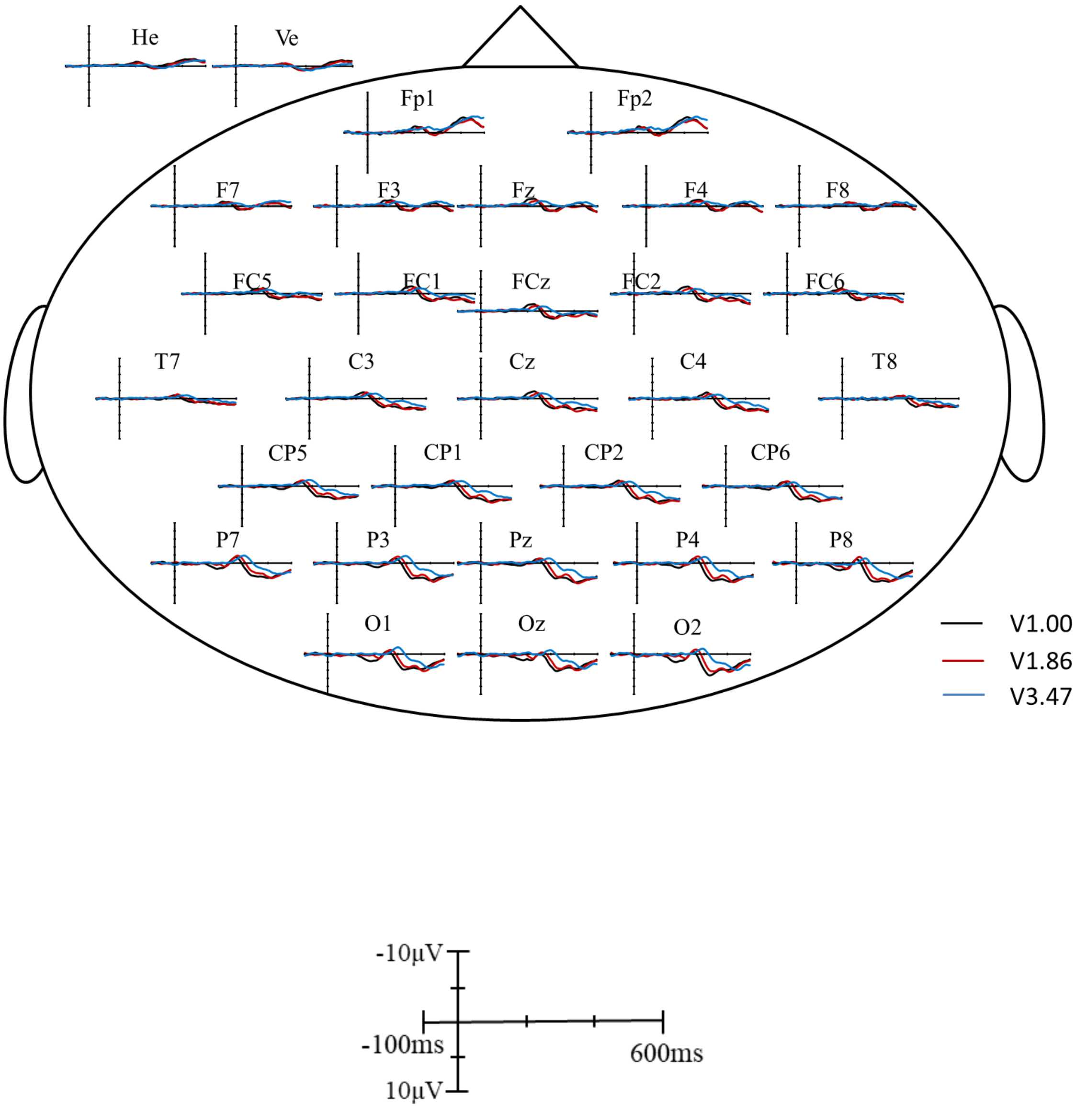


Figure AII-2. *Averaged Event-Related Potentials of visual stimulus with three spatial frequencies*

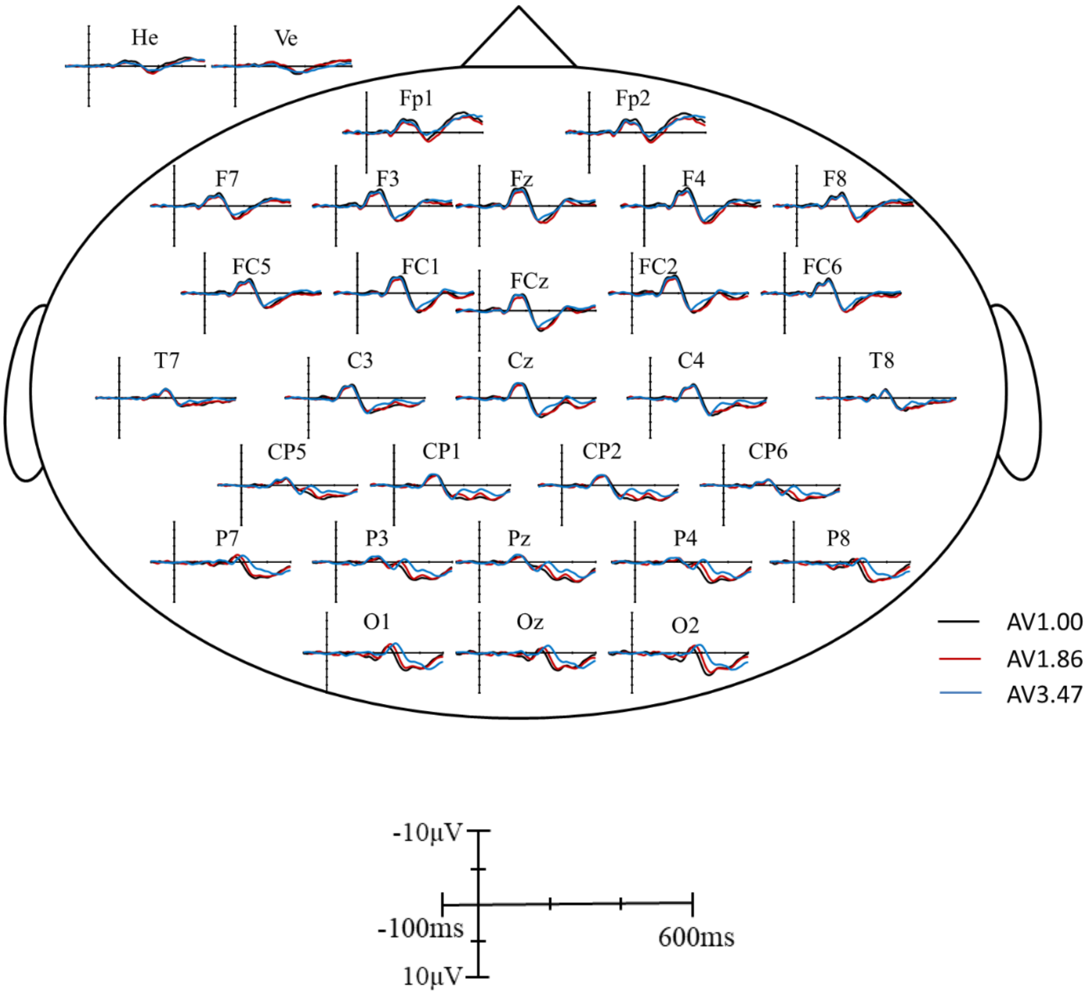


Figure AII-3. *Averaged Event-Related Potentials of audiovisual stimulus with three spatial frequencies*

Publications

1. Journal Paper

[1]. ***Fengxia Wu***, Yanna Ren, XiaoYu Tang, Qiong Wu, Yoshimichi Ejima, Jiajia Yang, Satoshi Takahashi, Jinglong Wu. Visual Identification Attenuates Audiovisual Interaction. Information, 2018, Vol. 21, No. 7.

[2]. Yanna Ren, Yanling Ren, Weiping Yang, Xiaoyu Tang, ***Fengxia Wu***, Qiong Wu, Satoshi Takahashi, Yoshimichi Ejima, Jinglong Wu. Comparison for younger and older adults: Stimulus temporal asynchrony modulates audiovisual integration. International Journal of Psychophysiology, 2018, Vol. 124, pp. 1-11.

[3]. Yanna Ren, Keisuke Suzuki, Weiping Yang, Yanling Ren, ***Fengxia Wu***, Jiajia Yang, Satoshi Takahashi, Yoshimichi Ejima, Jinglong Wu, and Koichi HirataYanna. Absent audiovisual integration elicited by peripheral stimuli in Parkinson's disease. Parkinson's Disease, 2018, Vol. 2018, No. 1648017.

2. International conference paper

[1]. ***Fengxia Wu***, Yanna Ren, Qiong Wu, Yoshimichi Ejima, Xiaoyu Tang, Weiping Yang, Jiajia Yang, Satoshi Takahashi, Jinglong Wu. Effects of Stimulus Features on Visual Processing for Communication between Human and Robot, IEEE International Conference on Mechatronics and Automation Harbin, 2017/08/07.

[2]. ***Fengxia Wu***, Yanna Ren, Xiaoyu Tang, Qiong Wu, Jiajia Yang, Weiping Yang, Satoshi Takahashi, Yoshimichi Ejima, Jinglong Wu. Attenuated Multisensory Integration in Cognitive Deficits Group, International Conference on Complex Medical Engineering (ICME), 2017:93.

[3]. ***Fengxia Wu***, Xiaoyu Tang, Weiping Yang, Yoshimichi Ejima, Qiong Wu, Yanna Ren, Takanori Ohara, Satoshi Takahashi, Jinglong Wu. Effects of Spatial Frequency on Audiovisual Integration for Communication between Human and Robot. International Conference on Mechatronics and Automation (ICMA). 2016: 1995-2000.

[4]. ***Fengxia Wu***, Miao Cao Weiping Yang, Yuki Ohara, Jinglong Wu. Effects of visual spatial frequency on audiovisual integration in a visual selective attention task. International Conference on Complex Medical Engineering (ICME), 2014:110.

[5]. Yanna Ren, Weiping Yang, Qiong Wu, ***Fengxia Wu***, Satoshi Takahashi, Yoshimichi Ejima, Jinglong Wu. Study of Audiovisual Asynchrony Signal Processing: Robot Recognition System of Different Ages. International Conference on Mechatronics and Automation (ICMA). 2016: 2320-2325.

[6]. Meng Wang, Qiong Wu, ***Fengxia Wu***, Jiajia Yang, Satoshi Takahashi, Yoshimichi

Ejima, Jinglong Wu. The time course of symmetry effect on shape perception: an event-related potential study. International Conference on Mechatronics and Automation (ICMA). 2018: 210-214.

3. Japan conference paper

[1]. Takanori Ohara, Satoshi Takahashi, Jiajia Yang, ***Fengxia Wu***, Yoshimichi Ejima, Jinglong Wu. Influence of auditory stimulus on perception of visual stimuli in different spatial frequencies. Annual Conference of the SICE Chugoku Chapter, 2015, pp.130-131.

4. Book chapter

[1]. ***Fengxia Wu***, Xiaoyu Tang, Yanna Ren, Weiping Yang, Satoshi Takahashi, Jinglong Wu. Effects of Visual Contrast on Inverse Effectiveness in Audiovisual Integration, Improving the Quality of Life for Dementia Patients through Progressive Detection, Treatment, and Care. Hershey, PA, USA: IGI Global; 2017. pp. 187-200.

[2]. Yanna Ren, Weiping Y, Xiaoyu T, ***Fengxia Wu***, Satoshi T, Jinglong W. The Early Diagnosis of Alzheimer's Disease: From Behavioral to Genetic Study. Improving the Quality of Life for Dementia Patients through Progressive Detection, Treatment, and Care. Hershey, PA, USA: IGI Global; 2017. pp. 1-16.

Acknowledgements

Firstly, I would like to express my sincerely gratitude to Prof. Jinglong Wu for the continuous support during my Ph.D studies and related researches. Prof. Jinglong Wu helped me in all my research design and writing of this thesis. I could not complete my study of doctor course and finish this thesis successfully without his enlightening instruction, impressive kindness and patience. His diligence gives me power not only during my present PhD scours, but also in my future life. In addition, Prof. Wu also helps me also for my daily life, and let my life in japan much easier.

Secondly, I would lie to express my sincerely gratitude to my supervisor Prof. Satoshi Takahashi. I got a lot of comments for Prof. Satoshi Takahashi during the research plan, conduct experiments, write published papers and this thesis. During the PhD scours, Prof. Satoshi Takahashi also greatly supports me when I applied the scholarships.

Thirdly, I also want to thank the rest of my thesis committee: Prof. Masanobu Abe, and Prof. Tokumi Yokohira for their insightful comments and encouragement, but also for the question which incented me to widen my research from various perspectives.

Last but not the least, I would also like to express my sincere thanks to Prof. Xiaoyu Tang and Lecturer Yanna Ren, who provided me a lot of comments during I write my papers. Without their precious support, it would not be possible to conduct this thesis successfully. I express sincere thanks to Assistant Professor Jiajia Yang. I thank my fellow lab mates in for the stimulating discussions and for all the fun we have had together during these years. I would like to thank my parents, my brother and sister and friends for supporting me spiritually throughout writing this thesis and my life in general.

References

[1] K. Grill-Spector and N. Kanwisher, "Visual recognition: As soon as you know it is there, you know what it is," *Psychological Science,* vol. 16, pp. 152-160, 2005.

[2] M. L. Mack, I. Gauthier, J. Sadr, and T. J. Palmeri, "Object detection and basic-level categorization: Sometimes you know it is there before you know what it is," *Psychonomic Bulletin & Review,* vol. 15, pp. 28-35, 2008.

[3] S. Straube and M. Fahle, "Visual detection and identification are not the same: Evidence from psychophysics and fMRI," *Brain and cognition,* vol. 75, pp. 29-38, 2011.

[4] B. Windey, W. Gevers, and A. Cleeremans, "Subjective visibility depends on level of processing," *Cognition,* vol. 129, pp. 404-409, 2013.

[5] M. Koivisto, S. Grassini, N. Salminen-Vaparanta, and A. Revonsuo, "Different electrophysiological correlates of visual awareness for detection and identification," *Journal of cognitive neuroscience,* vol. 29, pp. 1621-1631, 2017.

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

[7] M. S. Beauchamp, "See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex," *Curr Opin Neurobiol,* vol. 15, pp. 145-53, Apr 2005.

[8] B. E. Stein, "Development and organization of multimodal representation in cat superior colliculus," *Fed Proc,* vol. 37, pp. 2240-5, Jul 1978.

[9] T. M. Wright, K. A. Pelphrey, T. Allison, M. J. McKeown, and G. McCarthy, "Polysensory interactions along lateral temporal regions evoked by audiovisual speech," *Cerebral Cortex,* vol. 13, pp. 1034-1043, Oct 2003.

[10] C. Perrodin, C. Kayser, N. K. Logothetis, and C. I. Petkov, "Natural asynchronies in audiovisual communication signals regulate neuronal multisensory interactions in voice-sensitive cortex," *Proceedings of the National Academy of Sciences,* vol. 112, pp. 273-278, 2015.

[11] D. A. Slutsky and G. H. Recanzone, "Temporal and spatial dependency of the ventriloquism effect," *Neuroreport,* vol. 12, pp. 7-10, Jan 22 2001.

[12] J. F. Linden, A. Grunewald, and R. A. Andersen, "Responses to auditory stimuli in macaque lateral intraparietal area II. Behavioral modulation," *Journal of Neurophysiology,* vol. 82, pp. 343-358, 1999.

[13] M. A. Meredith and B. E. Stein, "Spatial determinants of multisensory integration in cat superior colliculus neurons," *Journal of Neurophysiology,* 1996.

[14] M. A. Meredith, J. W. Nemitz, and B. E. Stein, "Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors," *The Journal of*

neuroscience, vol. 7, pp. 3215-3229, 1987.

[15] J. M. Stephen, J. E. Knoefel, J. Adair, B. Hart, and C. J. Aine, "Aging-related changes in auditory and visual integration measured with MEG," *Neuroscience letters,* vol. 484, pp. 76-80, 2010.

[16] G. Hein and R. T. Knight, "Superior temporal sulcus—it's my area: or is it?," *Journal of Cognitive Neuroscience,* vol. 20, pp. 2125-2136, 2008.

[17] R. A. Stevenson and T. W. James, "Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition," *Neuroimage,* vol. 44, pp. 1210-23, Feb 1 2009.

[18] N. E. Barraclough, D. K. Xiao, C. I. Baker, M. W. Oram, and D. I. Perrett, "Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions," *Journal of Cognitive Neuroscience,* vol. 17, pp. 377-391, Mar 2005.

[19] R. A. Stevenson, M. L. Geoghegan, and T. W. James, "Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects," *Experimental Brain Research,* vol. 179, pp. 85-95, May 2007.

[20] J. A. Johnson, A. P. Strafella, and R. J. Zatorre, "The role of the dorsolateral prefrontal cortex in bimodal divided attention: two transcranial magnetic stimulation studies," *Journal of Cognitive Neuroscience,* vol. 19, pp. 907-920, 2007.

[21] J. A. Johnson and R. J. Zatorre, "Attention to simultaneous unrelated auditory and visual events: behavioral and neural correlates," *Cerebral Cortex,* vol. 15, pp. 1609-1620, 2005.

[22] L. M. Romanski, "Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex," *Cerebral Cortex,* vol. 17, pp. I61-I69, Sep 2007.

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

[24] D. Talsma, T. J. Doty, and M. G. Woldorff, "Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration?," *Cerebral cortex,* vol. 17, pp. 679-690, 2007.

[25] R. L. De Valois, E. W. Yund, and N. Hepler, "The orientation and direction selectivity of cells in macaque visual cortex," *Vision research,* vol. 22, pp. 531-544, 1982.

[26] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of physiology,* vol. 160, pp. 106-154, 1962.

[27] F. Campbell, J. Kulikowski, and J. Levinson, "The effect of orientation on the visual resolution of gratings," *The Journal of physiology,* vol. 187, p. 427, 1966.

[28] S. K. Petruno, R. E. Clark, and P. Reinagel, "Evidence that primary visual cortex is required for image, orientation, and motion discrimination by rats," *PLoS One,* vol. 8, p. e56543, 2013.

[29] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature*

reviews neuroscience, vol. 2, pp. 194-203, 2001.

[30] A. D. Milner and M. A. Goodale, "Two visual systems re-viewed," *Neuropsychologia,* vol. 46, pp. 774-785, 2008.

[31] P. Dean, "Grating detection and visual acuity after lesions of striate cortex in hooded rats," *Experimental brain research,* vol. 43, pp. 145-153, 1981.

[32] L. L. Glickfeld, M. H. Histed, and J. H. Maunsell, "Mouse primary visual cortex is used to detect both orientation and contrast changes," *Journal of Neuroscience,* vol. 33, pp. 19416-19422, 2013.

[33] S. Spotorno, G. S. Masson, and A. Montagnini, "Fixational saccades during grating detection and discrimination," *Vision research,* vol. 118, pp. 105-118, 2016.

[34] B. G. Breitmeyer, "Simple reaction time as a measure of the temporal response properties of transient and sustained channels," *Vision research,* vol. 15, pp. 1411-1412, 1975.

[35] S. Plainis and I. Murray, "Neurophysiological interpretation of human visual reaction times: effect of contrast, spatial frequency and luminance," *Neuropsychologia,* vol. 38, pp. 1555-1564, 2000.

[36] I. Murray and S. Plainis, "Contrast coding and magno/parvo segregation revealed in reaction time studies," *Vision research,* vol. 43, pp. 2707-2719, 2003.

[37] G. Sotiropoulos, A. R. Seitz, and P. Seriès, "Contrast dependency and prior expectations in human speed perception," *Vision research,* vol. 97, pp. 16-23, 2014.

[38] D. Sagi and B. Julesz, "Detection versus discrimination of visual orientation," *Perception,* vol. 13, pp. 619-628, 1984.

[39] J. Nachmias and R. V. Sansbury, "Grating contrast: discrimination may be better than detection," *Vision research,* vol. 14, pp. 1039-1042, 1974.

[40] Y. Ren, W. Yang, K. Nakahashi, S. Takahashi, and J. Wu, "Audiovisual Integration Delayed by Stimulus Onset Asynchrony Between Auditory and Visual Stimuli in Older Adults," *Perception,* p. 0301006616673850, 2016.

[41] E. Dzhafarov and M. Rauk, "Position discrimination may be better than detection," *Vision research,* vol. 22, pp. 1079-1081, 1982.

[42] B. R. Beutter, M. P. Eckstein, and L. S. Stone, "Saccadic and perceptual performance in visual search tasks. I. Contrast detection and discrimination," *JOSA A,* vol. 20, pp. 1341-1355, 2003.

[43] G.-J. Pepping and F.-X. Li, "Effects of response task on reaction time and the detection of affordances," *Motor Control,* vol. 9, pp. 129-143, 2005.

[44] C. D. Gilbert and W. Li, "Top-down influences on visual processing," *Nature Reviews Neuroscience,* vol. 14, pp. 350-363, 2013.

[45] L. S. Stone and P. Thompson, "Human speed perception is contrast dependent," *Vision research,* vol. 32, pp. 1535-1549, 1992.

[46] A. Felipe and M. J. Buades, "Influence of the contrast sensitivity function on the reaction time," *Vision Research,* vol. 33, pp. 2461-2466, 1993.

[47] R. S. Harwerth and D. M. Levi, "Reaction time as a measure of suprathreshold grating detection," *Vision research,* vol. 18, pp. 1579-1586, 1978.

[48] S. Braeutigam, A. J. Bailey, and S. J. Swithenby, "Task-dependent early latency (30–60 ms) visual processing of human faces and other objects," *Neuroreport,* vol. 12, pp. 1531-1536, 2001.

[49] R. Jones and M. J. Keck, "Visual evoked response as a function of grating spatial frequency," *Investigative ophthalmology & visual science,* vol. 17, pp. 652-659, 1978.

[50] L. Huang and K. R. Dobkins, "Attentional effects on contrast discrimination in humans: evidence for both contrast gain and response gain," *Vision research,* vol. 45, pp. 1201-1212, 2005.

[51] D. G. Pelli, "Uncertainty explains many aspects of visual contrast detection and discrimination," *JOSA A,* vol. 2, pp. 1508-1532, 1985.

[52] E. Orchard-Mills, E. Van der Burg, and D. Alais, "Amplitude-modulated auditory stimuli influence selection of visual spatial frequencies," *Journal of vision,* vol. 13, pp. 6-6, 2013.

[53] K. Brooks, "Stereomotion speed perception is contrast dependent," *Perception,* vol. 30, pp. 725-731, 2001.

[54] A. Vassilev and D. Mitov, "Perception time and spatial frequency," *Vision research,* vol. 16, pp. 89-92, 1976.

[55] J. J. McDonald, W. A. Teder-Sälejärvi, and S. A. Hillyard, "Involuntary orienting to sound improves visual perception," *Nature,* vol. 407, pp. 906-908, 2000.

[56] M. A. Meredith and B. E. Stein, "Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration," *Journal of neurophysiology,* vol. 56, pp. 640-662, 1986.

[57] B. E. Stein, N. London, L. K. Wilkinson, and D. D. Price, "Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis," *Journal of cognitive neuroscience,* vol. 8, pp. 497-506, 1996.

[58] M. Lippert, N. K. Logothetis, and C. Kayser, "Improvement of visual contrast detection by a simultaneous sound," *Brain research,* vol. 1173, pp. 102-109, 2007.

[59] F. Leo, V. Romei, E. Freeman, E. Ladavas, and J. Driver, "Looming sounds enhance orientation sensitivity for visual stimuli on the same side as such sounds," *Experimental brain research,* vol. 213, pp. 193-201, 2011.

[60] J. Vroomen and B. d. Gelder, "Sound enhances visual perception: cross-modal effects of auditory organization on vision," *Journal of experimental psychology: Human perception and performance,* vol. 26, p. 1583, 2000.

[61] Q. Li, H. Yang, F. Sun, and J. Wu, "Spatiotemporal relationships among audiovisual stimuli modulate auditory facilitation of visual target discrimination," *Perception,* vol. 44, pp. 232-242, 2015.

[62] Y.-C. Chen, P.-C. Huang, S.-L. Yeh, and C. Spence, "Synchronous sounds enhance visual sensitivity without reducing target uncertainty," *Seeing and perceiving,* vol. 24, pp. 623-638, 2011.

[63] R. Cecere, V. Romei, C. Bertini, and E. Làdavas, "Crossmodal enhancement of visual orientation discrimination by looming sounds requires functional

activation of primary visual areas: a case study," *Neuropsychologia,* vol. 56, pp. 350-358, 2014.
[64] S. Gleiss and C. Kayser, "Eccentricity dependent auditory enhancement of visual stimulus detection but not discrimination," *Frontiers in integrative neuroscience,* vol. 7, p. 52, 2013.
[65] S. J. Kayser, M. G. Philiastides, and C. Kayser, "Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations," *NeuroImage,* 2017.
[66] S. J. Luck and S. A. H. Yard, "The role of attention in feature detection and conjunction discrimination: An electrophysiological analysis," *International Journal of Neuroscience,* vol. 80, pp. 281-297, 1995.
[67] D. H. Brainard and S. Vision, "The psychophysics toolbox," *Spatial vision,* vol. 10, pp. 433-436, 1997.
[68] R. A. Stevenson, D. Ghose, J. K. Fister, D. K. Sarko, N. A. Altieri, A. R. Nidiffer, *et al.*, "Identifying and quantifying multisensory integration: a tutorial review," *Brain topography,* vol. 27, pp. 707-730, 2014.
[69] F. Frassinetti, N. Bolognini, and E. Làdavas, "Enhancement of visual perception by crossmodal visuo-auditory interaction," *Experimental brain research,* vol. 147, pp. 332-343, 2002.
[70] T. Noesselt, S. Tyll, C. N. Boehler, E. Budinger, H.-J. Heinze, and J. Driver, "Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity," *Journal of Neuroscience,* vol. 30, pp. 13609-13623, 2010.
[71] D. Senkowski, D. Saint-Amour, M. Höfle, and J. J. Foxe, "Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness," *Neuroimage,* vol. 56, pp. 2200-2208, 2011.
[72] V. Romei, M. M. Murray, C. Cappe, and G. Thut, "The contributions of sensory dominance and attentional bias to cross-modal enhancement of visual cortex excitability," *Journal of Cognitive Neuroscience,* vol. 25, pp. 1122-1135, 2013.
[73] A. Falchier, S. Clavagnier, P. Barone, and H. Kennedy, "Anatomical evidence of multimodal integration in primate striate cortex," *Journal of Neuroscience,* vol. 22, pp. 5749-5759, 2002.
[74] A. L. Beer, T. Plank, and M. W. Greenlee, "Diffusion tensor imaging shows white matter tracts between human auditory and visual cortex," *Experimental Brain Research,* vol. 213, p. 299, 2011.
[75] P. J. Laurienti, M. T. Wallace, J. A. Maldjian, C. M. Susi, B. E. Stein, and J. H. Burdette, "Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices," *Human brain mapping,* vol. 19, pp. 213-223, 2003.
[76] J. L. Mozolic, C. E. Hugenschmidt, A. M. Peiffer, and P. J. Laurienti, "Modality-specific selective attention attenuates multisensory integration," *Experimental brain research,* vol. 184, pp. 39-52, 2008.
[77] M. C. Potter, M. M. Chun, B. S. Banks, and M. Muckenhoupt, "Two attentional deficits in serial target search: the visual attentional blink and an amodal task-

switch deficit," *Journal of Experimental Psychology: Learning, Memory, and Cognition,* vol. 24, p. 979, 1998.
[78] J.-i. Kawahara, V. Di Lollo, and J. T. Enns, "Attentional requirements in visual detection and identification: Evidence from the attentional blink," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 27, p. 969, 2001.
[79] K. D. Gibney, E. Aligbe, B. A. Eggleston, S. R. Nunes, W. G. Kerkhoff, C. L. Dean*, et al.*, "Visual distractors disrupt audiovisual integration regardless of stimulus complexity," *Frontiers in integrative neuroscience,* vol. 11, p. 1, 2017.
[80] A. Pérez-Bellido, S. Soto-Faraco, and J. López-Moliner, "Sound-driven enhancement of vision: disentangling detection-level from decision-level contributions," *Journal of neurophysiology,* vol. 109, pp. 1065-1077, 2013.
[81] R. De Meo, M. M. Murray, S. Clarke, and P. J. Matusz, "Top-down control and early multisensory processes: chicken vs. egg," *Frontiers in integrative neuroscience,* vol. 9, p. 17, 2015.
[82] B. E. Stein, M. A. Meredith, and M. T. Wallace, "The visually responsive neuron and beyond: multisensory integration in cat and monkey," *Prog Brain Res,* vol. 95, pp. 79-90, 1993.
[83] J. Krueger Fister, R. A. Stevenson, A. R. Nidiffer, Z. P. Barnett, and M. T. Wallace, "Stimulus intensity modulates multisensory temporal processing," *Neuropsychologia,* Feb 23 2016.
[84] W. Yang, J. Yang, Y. Gao, X. Tang, Y. Ren, S. Takahashi*, et al.*, "Effects of Sound Frequency on Audiovisual Integration: An Event-Related Potential Study," *PloS one,* vol. 10, p. e0138296, 2015.
[85] S. Rach, A. Diederich, and H. Colonius, "On quantifying multisensory interaction effects in reaction time and detection rate," *Psychological research,* vol. 75, pp. 77-94, 2011.
[86] M. Lippert, N. K. Logothetis, and C. Kayser, "Improvement of visual contrast detection by a simultaneous sound," *Brain Res,* vol. 1173, pp. 102-9, Oct 10 2007.
[87] T. Noesselt, S. Tyll, C. N. Boehler, E. Budinger, H.-J. Heinze, and J. Driver, "Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity," *The Journal of Neuroscience,* vol. 30, pp. 13609-13623, 2010.
[88] Y. Takeshima and J. Gyoba, "Low-Level Visual Processing Speed Modulates Judgment of Audio-Visual Simultaneity," *Interdisciplinary Information Sciences,* vol. 21, pp. 109-114, 2015.
[89] Y. Takeshima and J. Gyoba, "Spatial frequency modulates the degree of illusory second flash perception," *Multisensory research,* vol. 28, pp. 1-10, 2015.
[90] E. Orchard-Mills, E. Van der Burg, and D. Alais, "Amplitude-modulated auditory stimuli influence selection of visual spatial frequencies," *J Vis,* vol. 13, 2013.
[91] A. Perez-Bellido, S. Soto-Faraco, and J. Lopez-Moliner, "Sound-driven

enhancement of vision: disentangling detection-level from decision-level contributions," *J Neurophysiol,* vol. 109, pp. 1065-77, Feb 2013.

[92] B. E. Stein, N. London, L. K. Wilkinson, and D. D. Price, "Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis," *J Cogn Neurosci,* vol. 8, pp. 497-506, Nov 1996.

[93] J. J. McDonald, W. A. Teder-Salejarvi, and S. A. Hillyard, "Involuntary orienting to sound improves visual perception," *Nature,* vol. 407, pp. 906-8, Oct 19 2000.

[94] J. B. Levitt and J. S. Lund, "Contrast dependence of contextual effects in primate visual cortex," *Nature,* vol. 387, p. 73, 1997.

[95] B. C. Skottun, A. Bradley, G. Sclar, I. Ohzawa, and R. D. Freeman, "The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior," *Journal of neurophysiology,* vol. 57, pp. 773-786, 1987.

[96] R. L. De Valois, H. Morgan, and D. M. Snodderly, "Psychophysical studies of monkey vision-III. Spatial luminance contrast sensitivity tests of macaque and human observers," *Vision research,* vol. 14, pp. 75-81, 1974.

[97] L. Huang, "Visual Features: Featural Strength and Visual Strength Are Two Dissociable Dimensions," *Sci Rep,* vol. 5, p. 13769, 2015.

[98] B. E. Stein and M. A. Meredith, *The merging of the senses*: The MIT Press, 1993.

[99] B. E. Stein, M. A. Meredith, and M. T. Wallace, "The visually responsive neuron and beyond: multisensory integration in cat and monkey," *Progress in brain research,* vol. 95, pp. 79-90, 1993.

[100] P. M. Jaekl and S. Soto-Faraco, "Audiovisual contrast enhancement is articulated primarily via the M-pathway," *Brain research,* vol. 1366, pp. 85-92, 2010.

[101] Q. Li, H. Yu, X. Li, H. Sun, J. Yang, and C. Li, "The informativity of sound modulates crossmodal facilitation of visual discrimination: a fMRI study," *NeuroReport,* vol. 28, pp. 63-68, 2017.

[102] Q. Li, H. Yu, Y. Wu, and N. Gao, "The spatial reliability of task-irrelevant sounds modulates bimodal audiovisual integration: An event-related potential study," *Neuroscience Letters,* vol. 629, pp. 149-154, 2016.

[103] L. E. Leguire, A. Algaze, N. H. Kashou, J. Lewis, G. L. Rogers, and C. Roberts, "Relationship among fMRI, contrast sensitivity and visual acuity," *Brain research,* vol. 1367, pp. 162-169, 2011.

[104] P. Zhang, H. Zhou, W. Wen, and S. He, "Layer-specific response properties of the human lateral geniculate nucleus and superior colliculus," *Neuroimage,* vol. 111, pp. 159-166, 2015.

[105] T. U. Otto, B. Dassy, and P. Mamassian, "Principles of multisensory behavior," *Journal of Neuroscience,* vol. 33, pp. 7463-7474, 2013.

[106] C. Cappe, M. M. Murray, P. Barone, and E. M. Rouiller, "Multisensory facilitation of behavior in monkeys: effects of stimulus intensity," *Journal of Cognitive Neuroscience,* vol. 22, pp. 2850-2863, 2010.

[107] J. K. Siemann, C. L. Muller, G. Bamberger, J. D. Allison, J. Veenstra-

VanderWeele, and M. T. Wallace, "A novel behavioral paradigm to assess multisensory processing in mice," *Frontiers in behavioral neuroscience,* vol. 8, 2014.

[108] J. K. Bizley, R. K. Maddox, and A. K. Lee, "Defining auditory-visual objects: Behavioral tests and physiological mechanisms," *Trends in neurosciences,* vol. 39, pp. 74-85, 2016.

[109] H. Atilgan, S. Town, K. Wood, G. Jones, R. Maddox, A. Lee*, et al.*, "Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding," *bioRxiv,* p. 098798, 2017.

[110] G. Iurilli, D. Ghezzi, U. Olcese, G. Lassi, C. Nazzaro, R. Tonini*, et al.*, "Sound-driven synaptic inhibition in primary visual cortex," *Neuron,* vol. 73, pp. 814-828, 2012.

[111] L. Lemus, A. Hernández, R. Luna, A. Zainos, and R. Romo, "Do sensory cortices process more than one sensory modality during perceptual judgments?," *Neuron,* vol. 67, pp. 335-348, 2010.

[112] D. Kumpik, H. Roberts, A. King, and J. Bizley, "Visual sensitivity is a stronger determinant of illusory," 2014.

[113] A. R. Nidiffer, R. A. Stevenson, J. K. Fister, Z. P. Barnett, and M. T. Wallace, "Interactions between space and effectiveness in human multisensory performance," *Neuropsychologia,* 2016.

[114] N. Stoep, S. Stigchel, T. Nijboer, and M. Smagt, "Audiovisual integration in near and far space: effects of changes in distance and stimulus effectiveness," *Experimental brain research,* vol. 234, pp. 1175-1188, 2016.

[115] A. Diederich and H. Colonius, "Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time," *Perception & psychophysics,* vol. 66, pp. 1388-1404, 2004.

[116] R. A. Stevenson and T. W. James, "Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition," *Neuroimage,* vol. 44, pp. 1210-1223, 2009.

[117] S. Molholm, W. Ritter, M. M. Murray, D. C. Javitt, C. E. Schroeder, and J. J. Foxe, "Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study," *Cognitive brain research,* vol. 14, pp. 115-128, 2002.

[118] E. Jones and T. Powell, "An anatomical study of converging sensory pathways within the cerebral cortex of the monkey," *Brain,* vol. 93, pp. 793-820, 1970.

[119] M. S. Beauchamp, B. D. Argall, J. Bodurka, J. H. Duyn, and A. Martin, "Unraveling multisensory integration: patchy organization within human STS multisensory cortex," *Nature neuroscience,* vol. 7, p. 1190, 2004.

[120] J. Hocking and C. J. Price, "The role of the posterior superior temporal sulcus in audiovisual processing," *Cerebral Cortex,* vol. 18, pp. 2439-2449, 2008.

[121] B. L. Allman, R. E. Bittencourt-Navarrete, L. P. Keniston, A. E. Medina, M. Y. Wang, and M. A. Meredith, "Do cross-modal projections always result in multisensory integration?," *Cerebral Cortex,* vol. 18, pp. 2066-2076, 2008.

[122] A. L. Beer, T. Plank, G. Meyer, and M. W. Greenlee, "Combined diffusion-

weighted and functional magnetic resonance imaging reveals a temporal-occipital network involved in auditory-visual object processing," *Frontiers in integrative neuroscience,* vol. 7, p. 5, 2013.

[123] D. Talsma and M. G. Woldorff, "Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity," *Journal of cognitive neuroscience,* vol. 17, pp. 1098-1114, 2005.

[124] M. Georgeson and G. Sullivan, "Contrast constancy: deblurring in human vision by spatial frequency channels," *The Journal of Physiology,* vol. 252, pp. 627-656, 1975.

[125] G. A. Calvert, R. Campbell, and M. J. Brammer, "Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex," *Current biology,* vol. 10, pp. 649-657, 2000.

[126] H. Holle, J. Obleser, S.-A. Rueschemeyer, and T. C. Gunter, "Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions," *Neuroimage,* vol. 49, pp. 875-884, 2010.

[127] W. H. Merigan and J. H. Maunsell, "How parallel are the primate visual pathways?," *Annual review of neuroscience,* vol. 16, pp. 369-402, 1993.

[128] K. Nakamura, T. Oga, T. Okada, N. Sadato, Y. Takayama, T. Wydell*, et al.*, "Hemispheric asymmetry emerges at distinct parts of the occipitotemporal cortex for objects, logograms and phonograms: a functional MRI study," *Neuroimage,* vol. 28, pp. 521-528, 2005.

[129] T. Sugihara, M. D. Diltz, B. B. Averbeck, and L. M. Romanski, "Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex," *Journal of Neuroscience,* vol. 26, pp. 11138-11147, 2006.

[130] J. A. Johnson and R. J. Zatorre, "Neural substrates for dividing and focusing attention between simultaneous auditory and visual events," *Neuroimage,* vol. 31, pp. 1673-1681, 2006.

[131] T. Rohe and U. Noppeney, "Cortical hierarchies perform Bayesian causal inference in multisensory perception," *PLoS Biology,* vol. 13, p. e1002073, 2015.

[132] Y. Ren, Y. Ren, W. Yang, X. Tang, F. Wu, Q. Wu*, et al.*, "Comparison for younger and older adults: stimulus temporal asynchrony modulates audiovisual integration," *International Journal of Psychophysiology,* vol. 124, pp. 1-11, 2018.

[133] Y. Gao, Q. Li, W. Yang, J. Yang, X. Tang, and J. Wu, "Effects of ipsilateral and bilateral auditory stimuli on audiovisual integration: a behavioral and event-related potential study," *NeuroReport,* vol. 25, pp. 668-675, 2014.

[134] S. J. Kayser, M. G. Philiastides, and C. Kayser, "Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations," *NeuroImage,* vol. 148, pp. 31-41, 2017.