

**Studies on mechanisms of visual object
recognition by manipulating the context at
hierarchical visual processing stages**

2021, 9

Yiyang Yu

Graduate School of Natural Science and Technology
(Doctor's Course)

OKAYAMA UNIVERSITY

Abstract

We can effortlessly detect, classify and recognize objects by sight, touch or body sensory, even though each object is different in appearance. However, the mechanism underlying object recognition has not yet been fully understood. The mechanisms underlying object recognition can be investigated by comparing recognition successes with recognition failures. In the present study, we examined visual object recognition by manipulating object relations in multiple levels, such as flanking category, memory color, 3D depth scene and audio-visual crossmodality.

Firstly, we addressed the violation effect of symbol type under crowding in chapter 2 in which we examined the spatial context effect at the high level of the visual processing. In daily life, we are often in complex scenes. In such complex scenes, our visual system is limited by crowding and nearby objects inhibit our recognition of target objects. In order to explore the mechanism underlying the crowding effects at the high level stage which remain unclear, we investigated the interaction between target and flanker composing of three category pairs constructed from difference symbol types (number vs number, number vs letter and number vs symbol), for a brief and long exposure time in experiment 1. Our result of critical spacing showed that as the category's effect became stronger, the intensity of crowding was reduced in longer exposure time. Using the visual masking paradigm, we evaluated the category's effect in experiment 2. We proposed that the crowding at the high-level processing first increased until the peak value at about 145 ms and then decreased with SOA, suggesting

that the crowding at high-level is similar to middle level during a specific time window.

Secondly, we addressed the violation effects of memory color and 3D depth scene in chapter 3 in which we examined the temporal context effect at the high level of the visual processing. When the positions of objects in the scene are disrupted and placed in a cluttered form, the recognition speed and accuracy of target objects are reduced. In order to explore the mechanism underlying the violation effect, we investigated the influence of memory color, and 3d depth scene using the ERP and sLORETA method. The significant difference between valid and invalid stimuli was observed at 450ms in color condition and was observed at 200ms and 400ms in depth condition. The significant difference between valid and invalid stimuli was found in color (425-480ms, Sub-Gyral) and depth condition (180-250ms, IPL; 425-480ms, MFG) by using sLORETA analysis. The significant difference between valid and invalid stimuli was mediated at theta and beta band in color condition, alpha and beta band in depth condition by using time frequency analysis. The significant difference was mediated at inferior frontal gyrus (IFG) in theta band (4–7 Hz) and superior parietal lobule (SPL) in beta2 band (16.5–20 Hz) under color condition. Violation effect occurred in different time window and different brain areas among memory color and 3d depth scene, suggesting that the violation of color memory and 3d depth scene are mediated by different brain mechanism.

Finally, we addressed auditory-visual violation effect in chapter 4 in which we examined the temporal context effect at the high level of the auditory-visual crossmodal processing. Top-down modulation ability and cross-modal integration ability are both

the important cognitive basis of compensation hypothesis. Recent studies related to cross-modal sensory integration have shown that the sounds facilitate visual object processing. However, little is known about the role of top-down modulation in auditory-visual processing. In order to explore the mechanism underlying the crossmodal congruency effect, we investigated the influence of auditory-to-visual context using the ERP method and an auditory-to-visual priming paradigm. We examined the effect of naturalistic sound on visual object categorization in semantically auditory-to-visual congruent and incongruent conditions, we employed six animal sounds as prime stimuli and eighteen animal pictures (three pictures for each type of animal) as target stimuli in a visual object categorization task. Our result showed that the auditory to visual priming was similar to the visual to auditory priming about N400 effect, frontal lobe activation and the early stage gamma-band activity in the previous study. However, the auditory to visual priming was not similar to the visual to auditory priming about the occipital lobe activation, the significant differences of lateralization in left middle frontal gyrus (IMFG) and right superior frontal gyrus (rSFG), and the higher frequency-time window of gamma-band activity in the previous study. This finding is different from the finding for visual-to-auditory by Schneider et al. This implies that auditory-to-visual crossmodal processing is different from visual-to-auditory processing.

In addition, the N400 effect was strongly dependent on the difference of semantic size between prime-target pairs. The dependency could be accounted for by a hypothetical model in which the semantic information driven by the auditory prime and the information of the visual target may be integrated during visual object processing,

suggesting that the congruency or incongruency between an auditory prime and a visual target may be mediated by the interaction of bottom-up and top-down systems.

Table of Contents

Abstract.....	I
Chapter 1 Introduction.....	1
Summary.....	1
1.1 Context effect.....	3
1.2 Visual processing.....	8
1.3 Visual object recognition.....	10
1.4 Method.....	12
1.4.1 Violation method.....	12
1.4.2 Behavioral Methods.....	13
1.4.3 Electroencephalogram Methods.....	14
1.5 The contents of the dissertation.....	16
Chapter 2 The mechanism underlying the flanker effect at high level: category difference effect between number and non-number.....	17
Summary.....	17
2.1 Background.....	18
2.2 General methods.....	22
2.2.1 Participants.....	22
2.2.2 Material and stimuli.....	22
2.3 Experiment 1: Flanker category effect is affected by exposure time.....	25
2.3.1 Method.....	25
2.3.2 Results and discussion.....	25

2.4 Experiment 2 The temporal property of flanker category	29
2.4.1 Method	29
2.4.2 Results and discussion.....	30
2.5 Discussion.....	32
2.6 Conclusion.....	34
Chapter 3 A comparison of the mechanisms underlying the memory color violation and the spatial configuration violation.....	35
Summary.....	35
3.1 Background.....	37
3.2 Method.....	39
3.2.1 Participants.....	39
3.2.2 Stimuli	39
3.2.3 Procedure.....	40
3.2.4 EEG recordings	41
3.2.5 ERP analysis.....	42
3.2.6 Time-frequency analysis	43
3.2.7 Analysis of source estimations.....	44
3.3 Result.....	45
3.3.1 Results of the behavioral experiment.....	45
3.3.2 Results of ERP	46
3.3.3 Results of significant difference of ERPs	48
3.3.4 Source estimations of ERPs	49

3.3.5 Result of time frequency analysis	50
3.4 Discussion.....	56
3.4.1 Comparison to behavior data indicate different context processing	56
3.4.2 Comparison to validity-related ERP data indicate different context processing.....	57
3.4.3 Comparison to validity-related ERSP data indicate different context processing.....	58
3.4.4 The relativity of brain region in different context processing.....	61
3.5 Conclusion.....	63
Chapter 4 The mechanism underlying the interaction processing between audition and vision by violation method.....	64
Summary.....	64
4.1 Background.....	66
4.2 Methods	69
4.2.1 Participants	69
4.2.2 Stimuli	69
4.2.3 Procedure.....	71
4.2.4 Data recordings	74
4.2.5 Data analysis	74
4.3 Results	77
4.3.1 Behavioral data.....	77
4.3.2 Effects of semantic congruency on ERPs	77

4.3.3 Source estimations of ERPs	80
4.3.4 Effects of semantic congruency on gamma-band activity	82
4.3.5 Results of ERPs classified by prime-target pairs	84
4.4.6 Correlation between ERP amplitude difference and semantic size difference.....	85
4.4 Discussion.....	89
4.4.1 Multisensory effects on event-related potentials.....	89
4.4.2 Source localization of multisensory effects	91
4.4.3 Multisensory effects on gamma-band activity	92
4.4.4 Multisensory incongruency is affected by the expected semantic size	93
4.5 Conclusion	96
Chapter 5. Conclusion.....	97
Acknowledgments.....	99
Appendix.....	100
Reference	102

Chapter 1 Introduction

Summary

Visual object recognition is the ability of humans to identify target objects in complex scenes efficiently. Even if the visual message is incomplete or the context surrounding the visual target changes, humans are still able to recognize objects successfully. In real-world environments, objects usually do not appear independently, and the other object's presence can affect the recognition of the target. These effects may reduce the efficiency of target recognition or promote it. This effect is called context effect. Context effects may arise from objects which appear together, objects which appeared before, even rules in memory or cross-modal information.

This chapter introduces the concepts of context effect, visual processing and visual object recognition. In addition, the methods involved in this study are also presented. The relevant aspects of previous studies on the influence of context effects on visual object recognition are summarized in this chapter.

In Section 1, the concepts of temporal context effect and spatial context effect are introduced. In Section 2, the concepts of visual processing, high level vision and low level vision are introduced. In Section 3, the concept of visual target recognition and the processing hierarchy of visual target recognition are introduced. In Section 4, the impact of contextual effects on visual target recognition is introduced. In Section 5 the behavioral and EEG analysis methods used in the study are presented. In Section 6 the

CHAPTER 1 INTRODUCTION

research objectives of each section of the thesis are presented.

Our aim is to study the mechanisms underlying visual target recognition processes in different contexts using behavioral and EEG methods.

1.1 Context effect

‘No man is an island.’ That is, the perception of a target depends on its spatial context (features around the target object) and temporal context (sensory input in the most recent time). Numerous psychophysical studies have shown that contextual information affects visual, auditory, and somatosensory processing [1]. Visual spatial context effect has been studied extensively with a variety of stimuli, including simple visual features, such as a target (e.g. a circle, lines, gratings, number, faces et al.) surrounded by flanking stimuli. Examples of spatial context at the low, middle and high level visual processing stages are shown in Fig. 1.1.

Example 1~3 at the low level illustrate that the perceived brightness or color of a test circle or line is altered by the presence of surrounding circles with different luminance or color with different color [2].

Example 1 at the middle level illustrates that ability to recognize the misalignment between the two central lines (vernier acuity) depends on the flanker conditions [3]. The acuity improves as the flankers drift away from the target central lines. Example 2 at the middle level illustrates that the contrast threshold of Gabor patches in the center depends on test-flanker distance [4]. Example 3 at the middle level illustrates that the vertical test grating in the center appears repelled in orientation away from the surrounding grating (left and right) [5].

Example 1~2 at the high level that the identification of a central target number ‘3’ is more impaired by number flankers than by letter flankers [6]. The identifying proves as the flankers drift away from the target.

Example 3 at the high level that the identification of a central target face is crowded by normal face flankers more than inverted face flankers [7].

Previous reviews have extensively described low level features [8, 9], and there is little debate over the property of crowding for these high level features.

However, although a lot of evidence has been accumulated, it remains fully unestablished [10].

Manassi and Whitney have proposed the aims of the future crowding model: 1. How crowding occurs at several levels of visual processing; 2. How the information in crowded objects survives under different visual stages. [10].

Contrast


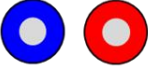




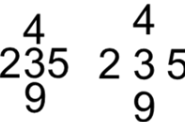
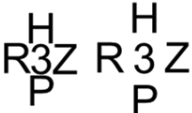

Visual processing level	Performance and effect	Example 1	Example 2	Example 3
Low level	Brightness contrast Color contrast	 Brightness contrast	 Color contrast	 Color contrast
Middle level	Vernier acuity Grating contrast Tilt illusion	 Vernier acuity	 Tilt illusion	 Grating contrast
High level (memory effect)	Flanking effect			

Fig. 1.1. Examples of spatial context effect at low, middle and high level. Examples of spatial context effect at low, middle and high level. Low level: Example1-3: The perceived brightness or color of a test circle or line is altered by the presence of surrounding circle or grating [2].

CHAPTER 1 INTRODUCTION

Middle level: Example 1: The ability to recognize the misalignment between the central lines (Vernier acuity) is different in the different flanker conditions. The acuity improves as the flankers drift away from the target central lines [3]. Example 2: Gabor patches of fixed contrast in the center appear differently in the flanker different conditions [4]. Example 3: The vertical test grating in the center appears repelled in orientation away from the surrounding grating (left and right) [5]. High level: Example 1~2: Identifying a central target number '3' is more impaired by flanking numbers than by flanking letters. The identifying proves as the flankers drift away from the target [6]. Example 3: Identifying a central target face is crowded by flanking upright faces more than flanking inverted faces [7].

On the other hand, examples of temporal contexts at low, middle and high level visual processing stages are shown in Fig.1.2. The example at the low level illustrates the perceived color of a white target is changed after prolonged inspection of color stimuli [11].

The example at the middle level illustrates that the perceived orientation of a vertical grating is changed after prolonged inspection of another oriented grating [12].

The example of a high level (same modality) illustrates that 'My wife and my mother in-law', a famous image [13], can be perceived as a young girl but not an old woman after hearing 'spoke a young girl word' as a cue. The example at high level (crossmodality) illustrates that unless one is already familiar with the picture such as hidden figure, dalmatian dog [14], it will not make much sense. There are just black patches on a white background. However, the word 'a dog' as a prime cue helps one to

recognize a dalmatian dog in the figure.

Previous reviews have extensively described the low level features, and there is little debate over the existence of adaptation effect for these types of stimuli [11, 15-17]. The prime cue effects have been reported by many previous researchers, however, the mechanism underlying the effects remains unclear.

In this study, we focus on the following three topics concerning spatial and temporal contexts in human vision.

Visual processing level	Performance	Example
Low level	Color adaptation	
Middle level	Orientation adaptation	
High level 1 (unimodality)	Cue effect on recognition	
High level 2 (crossmodality)	Crossmodal cue effect on recognition	

Fig. 1.2. Examples of temporal context effect at low, middle and high level. Low level: The perceived color of a white target is changed after prolonged inspection of color stimuli [11]. Middle level: The perceived orientation of a vertical grating is changed after prolonged inspection of another oriented grating [12]. High level(same modality): My wife and my mother in-law a famous image [13], which can be perceived either as a young girl or an old woman. However, the image can be perceived as a young girl but not an old woman after hearing ‘spoke

CHAPTER 1 INTRODUCTION

a young girl word' as a cue. High level (crossmodality): Unless one is already familiar with the picture such as hidden figure- dalmatian dog [14], it will not make much sense. There are just black patches on a white background. However, the word 'a dog' as a prime cue helps one to recognize a dalmatian dog in the figure.

1.2 Visual processing

Vision is the most important perception for humans, and we use it to obtain the most information. Visual processing begins with light to stimulate the retina and then projects through the lateral geniculate nucleus to the primary visual cortex (visual area 1, V1, BA 17 or striate cortex). Subsequently, the visual information is transmitted to the extrastriate areas. The extrastriate areas consist of visual areas 2, 3, 4, and 5 (V2, V3, V4, and V5, or BA 18 and BA 19) [18].

According to the two-streams hypothesis [19], the visual information is transmitted via the dorsal and ventral streams from V1. The ventral stream begins in V1, followed by V2, V4, and enters the inferior temporal lobe. Which is involved in the processing of object recognition. The dorsal stream starts from V1, passes through V2, enters the dorsomedial and temporal areas, and then reaches the parietal lobule. Which is involved in the processing of target spatial location information.

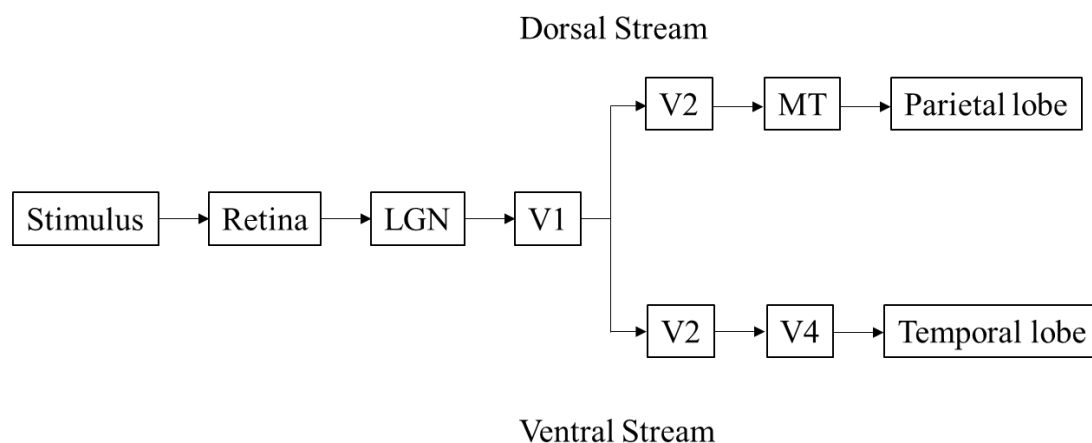


Fig. 1.3. Two-streams hypothesis of visual processing [20].

The hierarchical framework is typically used to explain visual processing, Low-level vision is processing information such as contrast, spatial frequency, and color. This is

the process of acquiring image attributes from the retina. Mid-level vision is processing shape features, texture and 3D depth information. This is the process of integrating image properties into the perceptual organization. High-level vision is responsible for face recognition, body movement recognition and visual target recognition. This is the process of applying perceptual organization to everyday life [21].

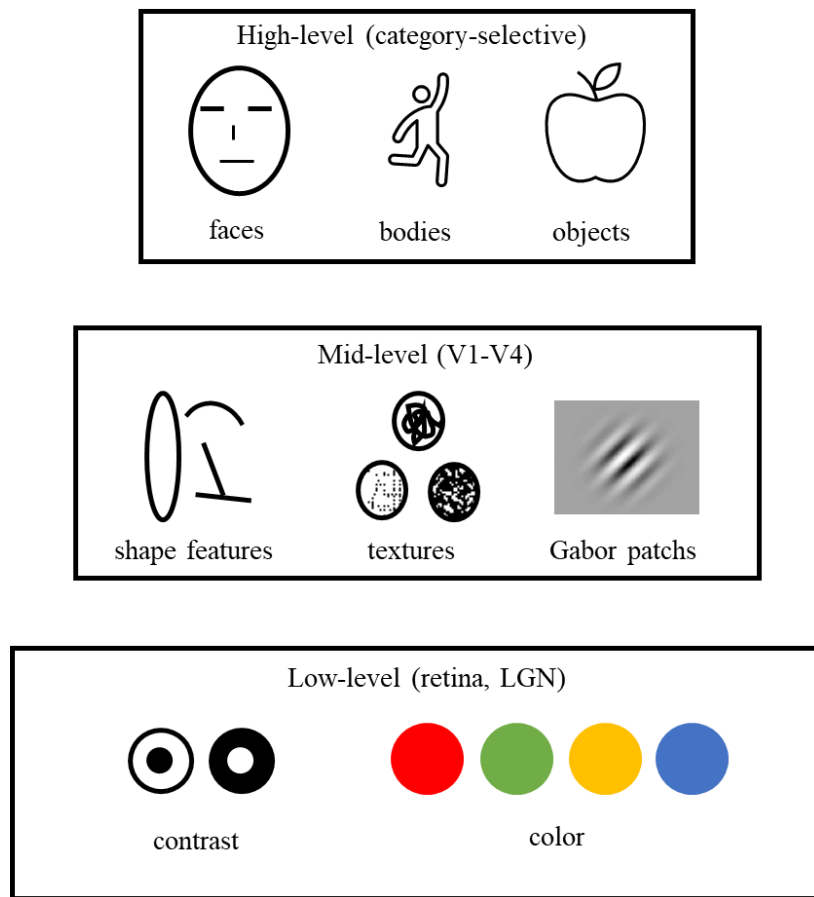


Fig. 1.4. Hierarchical framework of visual perception [22].

1.3 Visual object recognition

In our daily lives, we can recognize words in a book, birds in the sky, or cars on the road within a few hundred milliseconds. We can easily recognize these objects even if they are in different viewpoints, different illumination, or different contexts. This ability is called visual object recognition. By general definition, visual object recognition refers to the ability to quickly identify target objects in complex scenes, even if the scene is complex or the visual features of the target are obscured. This ability is important for most animals, including humans. Visual object recognition begins with the stimulation of the retina by light, the eye converts these stimuli into signals, and the brain processes this information in a visual representation subsequently. This process is so amazing that researchers in many fields have focused on research in this area. Cognitive neuroscience researchers prefer to study the activation areas of the brain and the time course of processing during visual object recognition; computational neuroscience researchers prefer to describe the process of visual object recognition using mathematical models, and computer vision researchers algorithmize the process of visual object recognition and execute it on a computer.

The core processing of visual object recognition is to match the visual information of the object with the relative concept in the brain. The hierarchical framework for visual processing is used to describe the process of visual target recognition because visual target recognition is highly relevant to high-level vision. It is widely accepted that the matching process is divided into four stages, as shown in Table 1 [23, 24].

Table 1.1. Stages of object recognition

Stage	Processing
Stage 1	Processing of basic object components.
Stage 2	Separating targets and background.
Stage 3	Matching the visual representation.
Stage 4	Using semantic information to promote recognition.

In the first stage, the basic components of vision are processed; in the second stage, the basic components are grouped; in the third stage, the visual representations and memory representations are matched; and in the fourth stage, semantic retrieval enables object classification and recognition.

1.4 Method

1.4.1 Violation method

The violation method which we use in Chapter 3 and Chapter 4 of this study is useful in research of cognitive processing in human vision. We describe the basis and its effectiveness below.

When objects are seen in an expected congruent context and at an expected position and size, object search and recognition are facilitated [25].

On the other hand, when objects are seen in an incongruent context or with unfamiliar distracter objects, visual performance is reduced because prediction or expectation is hindered. A yellow banana is recognized as a natural object, however, an unfamiliar color (e.g. blue) banana-shape object is not recognized as a natural object. It is effective to compare the visual performances for scene and object in an expected congruent condition and an unexpected incongruent conditions, by manipulating prediction or expectation. Indeed, comparing visual performances for a yellow banana and a blue banana-shape object is considered to be quite effective to examine the underlying mechanism of object recognition. In Chapter 3 and Chapter 4 of this study, we use a violation method that hinders prediction or expectation by adding unnatural properties to objects and scenes.

1.4.2 Behavioral Methods

Accuracy and reaction time

Accuracy and reaction time are important indexes in cognitive psychology experiments. Accuracy reflects the probability of participants' correct responses in the experiment, and reaction time reflects the corresponding speed of participants in the experiment. Accuracy and reaction time are influenced by cognitive ability and information processing performance [26].

Threshold and QUEST method

Humans derive their perception from the effects of stimuli. The threshold is the weakest stimulus intensity that humans can perceive. In psychophysics, researchers measure sensory thresholds to evaluate participants' ability to perceive stimuli under given conditions. There are many ways to measure sensory thresholds, and adaptive measurement methods are more popular as they can measure thresholds quickly and accurately. The QUEST method is an adaptive threshold measure using Bayesian methods [27]. We used the functions of psychtoolbox to implement this method [28]. The method is based on the general idea that the form of the human psychometric function is invariant when it is expressed as a logarithmic intensity function. [29]. Specifically, this method automatically adjusts the intensity parameters of the current stimulus based on the intensity of the previous stimulus and the participant's response. In short, the participant's response is correct, the difficulty increases, and the response

is incorrect, the difficulty decreases. Following this way, after a certain number of repetitions, we will obtain the participant's sensory threshold.

Visual masking method

Visual masking refers to the phenomenon that the target object recognition is more difficult when other objects appear. When the masking object is presented after the target appears, it is called backward masking. The phenomenon of visual masking is used to change the visibility of an object [30]. It is a typical method of visual research [31]. We can adjust the SOA between stimulus and masking to study the temporal properties of visual processing.

1.4.3 Electroencephalogram Methods

Electroencephalogram (EEG) is a technology that capturing the electrical signals on nerves in order to observe the human brain [32]. The Electrodes on the scalp are used to capture the tiny voltage waves caused by the firing of neurons in the brain. (Electrodes on the scalp are used to capture the tiny voltage waves caused by the firing of neurons in the brain.) EEG has a low spatial resolution but a high temporal resolution, and it is an effective tool to record brain activity.

Event-related potential (ERP) is the response of the brain caused by a specific sensory, cognitive or motor event [33]. In order to study the relationship between events and brain responses, the ERP method was proposed. Specific events elicit brain responses, but event-related features are hidden in many neural responses. To extract

event-related brain wave features, researchers average multiple EEG results to obtain relatively robust, well-defined waveforms. Such event-related brain waves are called ERPs. ERP is an effective tool for cognitive studies of temporal properties.

Special EEG frequencies imply special cognitive functions. Researchers use the Fourier transform to extract the features of EEG from different frequency bands. However, the results of frequency domain analysis do not contain temporal information, which makes the interpretation of the results difficult. Time-frequency analysis can retain both time-domain and frequency-domain features of the EEG to uncover more information. In psychological research, wavelet analysis is one of the most commonly used tools for time-frequency analysis [34]. The key idea of wavelet analysis is to select a function as the mother wavelet and generate a series of sub wavelets by transforming the mother wavelet. Different sub wavelets are used to respond to specific EEG frequencies, and then, the time-frequency features (energy, phase, etc.) of the EEG signal are extracted using convolution methods.

Standardized low-resolution brain electromagnetic tomography algorithm (sLORETA) is a method that uses EEG signals to identify areas of neural activity. It is one of the methods to solve the inverse problem of EEG. This method can generate normalized current density images with zero localization error. However, the spatial resolution of this method is relatively low [35]. In this study, the sLORETA software toolbox was used to preliminarily analyze the location of the origins of EEG features under specific conditions.

1.5 The contents of the dissertation

In chapter 1, the concept of visual processing, visual object recognition and context effect was introduced. The previous studies have also been summarized here. In addition, the methods of behavioral, ERP and sLORETA have been introduced. Lastly, the purpose of the thesis is emphasized.

In chapter 2, we addressed the violation effect of symbol type under crowding. According to the behavioral results, we proposed that the crowding at the high-level processing first increased until the peak value at about 145ms and then decreased with SOA, suggesting that the crowding at high-level is similar to the crowding at middle-level.

In chapter 3, we addressed the violation effects of memory color and spatial configuration. According to the results of ERP and sLORETA. We proposed that the violation effect occurred in different time window and different brain areas among memory color, and 3d depth scene, suggesting that the violation of color memory and 3d depth scene are mediated by different brain mechanism.

In chapter 4, we addressed auditory-visual violation effect. According to the results of ERP and sLORETA. We proposed that the auditory-to-visual crossmodal processing is different from the visual-to-auditory processing. For a further analysis, the N400 effect was strongly dependent on the difference of semantic size between prime-target pairs. Suggesting that the congruency or incongruency between an auditory prime and a visual target may be mediated by the interaction of bottom-up and top-down systems.

Chapter 2 The mechanism underlying the flanker effect at high level: category difference effect between number and non-number

Summary

When the human visual system processes complex scenes, crowding is considered a bottleneck in interpreting visual information. Several studies have reported that visual crowding seems to occur in the early stage of visual processing. After that, extensive evidence suggested that high-level visual information has a major effect on crowding, which showed substantial interest in the impact of high-level visual information on crowding. However, the temporal property of high-level crowding is not well understood. Here, we performed a series of behavioral experiments to show how flanker category information and exposure time affected crowding's intensity. We found that as the category's effect becomes stronger, the intensity of crowding will be reduced in longer exposure time. The peak of this effect occurs at about 145ms after the stimulus appears. We propose that exposure time may modulate high-level information processing depth, which can adjust multi-level crowding.

Keywords: Multi-Level Crowding, Flanker Category Effect, Temporal Property, Behavioral Experiment.

2.1 Background

Crowding is a visual perception phenomenon, which can break our identification ability when a target is surrounded by flankers in peripheral vision [36-39]. For example, it is difficult for people to glimpse a face in a crowd. In the same way, it is also difficult to read a message on the navigator when people concentrate on driving. In these examples, face and message in peripheral vision are more difficult to be identified because of the impact of crowding. In the real world, objects typically do not appear alone. Most of the objects are surrounded by other objects in the complex visual scenes. The crowding mechanism is the key point to know how humans can efficiently identify objects in complex scenes [40].

Scientists have been interested in the existence of crowding effect since the 1970s. From a spatial domain perspective, extensive research has shown that crowding is typically thought to manifest as a kind of contextual modulation caused by the inhibitory interaction between the target and the flankers. Following experimental research, researchers believe the reason for crowding is the failures of feature integration, segmentation, and coarse resolution of spatial attention. From a temporal domain perspective, it is widely accepted that crowding occurs in the identification stage of visual processing [41].

In some studies, crowding is thought to occur in the early stage, which means that the factors to affect crowding originate from the low-level visual information [42-44]. Nevertheless, there is extensive evidence that high-level visual information has a major effect on crowding.

Central to this debate is the role of high-level visual information. In the low-level crowding view, the existing body of vision research suggests that visual crowding is a bottom-up process [42-44]. When low-level features are combined, crowding destroys the integrity of information and reduces the utilization rate of high-level information, which leads to difficulties in identification. In the High-level crowding view, the researchers suggest that crowding may involve multiple levels of visual processing [10]. The property of high-level concepts may counteract the crowding. There are many kinds of high-level information that will not be affected by crowding, and high-level information can affect crowding at several stages, which shows the importance of top-down process. To date, there has been little agreement regarding the impact of high-level visual information on crowding. Accordingly, the mechanism of multi-level crowding has become a hot research topic in recent years.

Researchers adjust the target-flanker similarity to explore the properties of crowding in previous studies. In studies of multi-level crowding, the target-flanker difference usually originates from high-level visual information. These differences include the following levels, such as shape composition, integrity, object configuration, and dynamic configuration. There is much evidence to show that the property of high-level information plays a vital role in order to know the mechanism of multi-level crowding. Researchers used several identification tasks with letter, number, or symbol flanker. They found that the category property of flankers can adjust the intensity of crowding. Moreover, in order to explain this observation, the flanker category effect was proposed. In this study, the letter flankers reduced the crowding intensity in the number

identification task, even if the visual characteristics of letters and numbers have been unified. Flanker category effect is vital in multi-level crowding research, and the paradigm provided a way to study the function of high-level information above low-level visual characteristics under crowding [6].

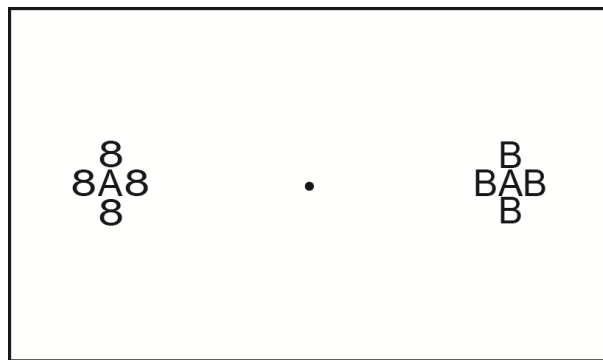


Fig. 2.1. Demonstration of flanker category effect. Left stimulus: examples of flanker category effect. Right stimulus: examples of normal crowding.

In support of low-level crowding view, the importance of temporal factor is widely studied [45-48]. Despite the importance of temporal factor, there remains a paucity of evidence on the role of temporal factor in multi-level crowding. According to the object recognition theory, the processing time will affect the depth of target processing, and the processing depth is an important factor to adjust the influence of the high-level information. Processing time might be an important factor that can affect the occurrence of flanker category effect.

This paper aims to prove the importance of processing time on flanking category effect. The central question in this dissertation asks how to determine the time frame of flanker category effect. The behavioral experiments were used to explore the flanker

CHAPTER 2 THE FLANKER CATEGORY EFFECT AT HIGH LEVEL

category effect in different processing times. This investigation used two experiments to achieve research purposes.

In Experiment 1, by using the QUEST method [27], the critical spacing under different temporal and category conditions was calculated. We used psychophysical methods to prove that the flanker category effect is related to temporal factors. In Experiment 2, we discussed the relationship between flanker category effect and processing time. The threshold of critical spacing is replaced by accuracy in this experiment because we want to investigate the results in strong crowding. Further, the visual masking method was used in this experiment to ensure accurate processing time. We adjusted the interstimulus interval between the stimulus and the masking to discuss the effect of processing time with short and long exposure time. This investigation will enhance our understanding of the role of processing time in multi-level crowding and provided a foundation for imaging and electrophysiological studies.

2.2 General methods

First, we will introduce the methodological information commonly used in all experiments.

2.2.1 Participants

Sixteen participants aged 24 to 33 years took part in this study. All had a normal or corrected-to-normal vision. Ten participated in Experiment 1, six in Experiment 2. All observers signed informed consent. Ethics was granted from Okayama University.

2.2.2 Material and stimuli

The experimental program is built by PTB based on MATLAB 2014a and presented on a Display++ LCD monitor (Cambridge Research System) with a frame rate of 100 Hz and resolution of 1920×1080 pixels [49]. Viewing distance was set to 57 cm secured using a chinrest. Eye movements were monitored using LiveTrack Lightning. This device is used to ensure that the observers are always looking at the center of the screen. Stimuli were black ($L_v^{1/4}$ 0.25 cd/m²) on white background ($L_v^{1/4}$ 91.5 cd/m²) with a high contrast ($C_{\text{michelson}^{1/4}}$ 0.99). A black fixation (0.3 deg diameter) was located in the center of the screen. The target was presented 15 deg from the fixation on the horizontal meridian randomly either in the left or the right visual field in Experiment 1. Moreover, the target was presented 10 deg from the fixation on the horizontal meridian randomly either on the left or the right in Experiment 2.

The experiment was started with a keypress. A fixation mark appeared at the center

CHAPTER 2 THE FLANKER CATEGORY EFFECT AT HIGH LEVEL

of the screen and stayed on the throughout block. Target and flankers were randomly displayed in the left or right visual field after varying onset times (0.9–1.5 s after trial onset) for 50 or 150ms, ensuring that no eye movements could be made. After the stimuli disappeared, all possible target stimuli were displayed on the screen for keyboard response. The next trial started after the observer's response, and the program provided audio feedback.

The font, which previous researchers designed, consisted of only seven straight lines like a calculator font. The targets were always numbers, 0–3. The letters J and Y and the numbers 7 and 9 were served as flankers since they shared the same features (in the designed font). Note that 9 and Y did not share the exact same feature set but had the same number of vertical and horizontal elements, with one horizontal element displaced in position. Meaningless flankers (symbols) were rotations or mirror images of the used letter flankers [6]. The size of visual stimuli in all experiments was 2.1 deg. The spacing between the target and the flanker is adaptive, and the value is determined by the quest toolbox in experiment 1.

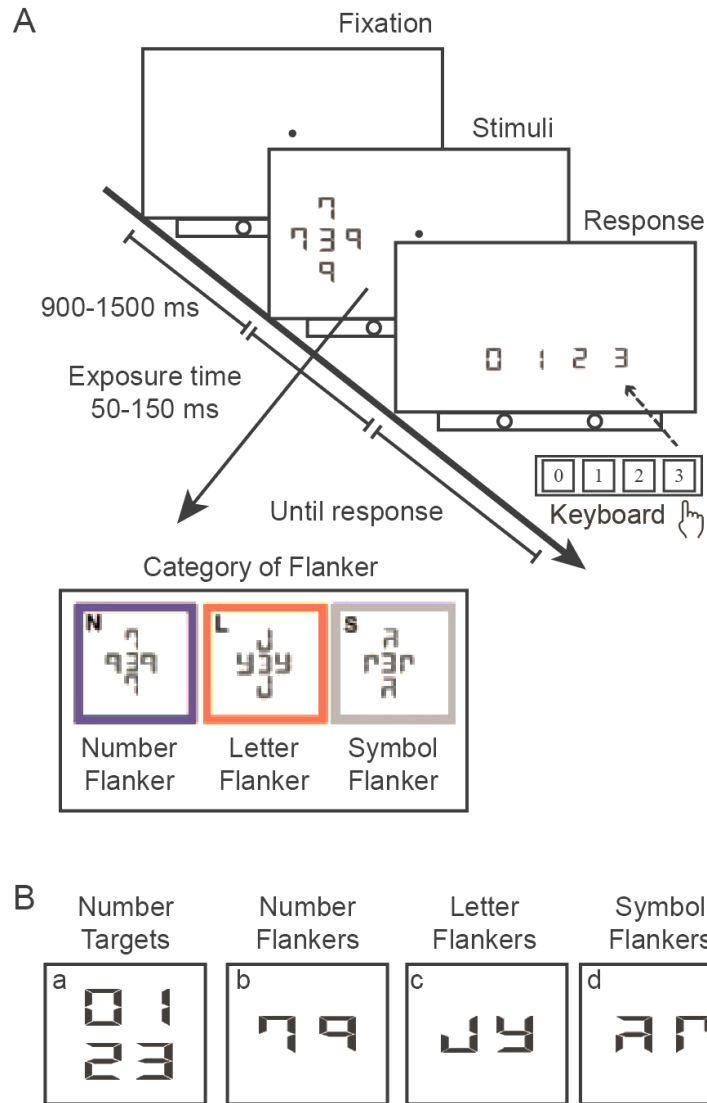


Fig 2.2 Experiment procedure and stimuli. The general procedure is shown in Fig 2.2A. Several target surround by four flankers is shown randomly in the right or left on the screen. Exposure time was changed in different experiments. The flanker type conditions are the same in all experiments, including number flanker, letter flanker, and symbol flanker. Fig 2.2B shows an example of the experimental stimulation used for all experiments. (a) are the number targets; (b) are the number flankers; (c) are the letter flankers. (d) are the symbol flankers.

2.3 Experiment 1: Flanker category effect is affected by exposure time

Experiment 1 was aimed to extend the results of previous study [6]. We extended that approach with a two (exposure time: 50 ms, 150 ms) by three (flanker: numbers, letters, and symbols) factorial design. To test that exposure time was involved in the effect of flanker category or not. This introduced the temporal property into the flanker category effect for the first time. If we can adjust the intensity of the category effect by manipulating the exposure time, it shows that the category effect is a dynamic process related to the processing depth of the target and the flanker.

2.3.1 Method

The QUEST algorithm was used to determine the critical spacing when the recognition accuracy rate is 82% [27]. The critical spacing was used as a measure of crowding intensity. Six blocks were classified according to conditions. Each block included less than 50 trials (If the participant's sight moves, the target number will be reselected for testing.), and it was decided to use the QUEST algorithm. The experiment was repeated five times, and the data were averaged, this method is same as previous study [6].

2.3.2 Results and discussion

The performance of the task was calculated by the degree of critical spacing. The results are presented in Fig. 2.3A and Fig. 2.3B. A two-way repeated measure analysis

CHAPTER 2 THE FLANKER CATEGORY EFFECT AT HIGH LEVEL

of variance (ANOVA) uncovered a significant effect of exposure time [$F(1, 9) = 8.984$ $P=.015$] and the type of flankers [$F(1.708, 15.375) = 8.239$ $P=.005$]. There is a significant interaction effect between exposure time and type of flankers [$F(1.600, 14.402) = 9.363$ $P=.004$]. Greenhouse–Geisser corrections were applied. Simple effects analyses were performed to uncover the relationship between two factors. The results are given in Table 1. Pairwise t-tests showed that longer exposure time could significantly reduce crowding under the condition of letter flanker. The effect of exposure time was not significant under the other two conditions. When the exposure time is 150ms, all the experimental conditions are consistent with the previous study, and we obtained consistent results with previous researchers. When the exposure time becomes shorter, the crowding intensity of letter flankers is significantly higher than that of number flankers, and the crowding intensity of letter flankers is the same as that of symbol flankers.

As mentioned in the previous study, flanker category effect was significant on the condition of 150 ms exposure time. This result suggests that crowding occurs at the object classification stage. However, when the exposure time was reduced, the observers could not complete the classification of letter flankers. This result may have led to an increase in the intensity of crowding. However, according to the results of this experiment, the trend of flanker category effect changing with the exposure time cannot be understood. For exposure time, we have defined only two levels. The current experimental design cannot rule out the possibility that the significant effect is simply caused by the change of difficulty.

Table 2.1 Results of simple effects analysis.

Pair of factors	Paired differences	t value	Sig.
NF 50 – NF 150	1.25	1.99	0.08
LF 50 – LF 150	2.98	4.37	0.01*
SF 50 – SF 150	1.27	1.84	0.10
NF 50 – LF 50	-1.39	-3.00	0.02*
NF 50 – SF 50	-1.23	-3.73	0.01*
LF 50 – SF 50	0.16	0.40	0.70
NF 150 – LF 150	0.34	1.28	0.23
NF 150 – SF 150	-1.21	-3.63	0.01*
LF 150 - SF 150	-1.55	-3.50	0.01*

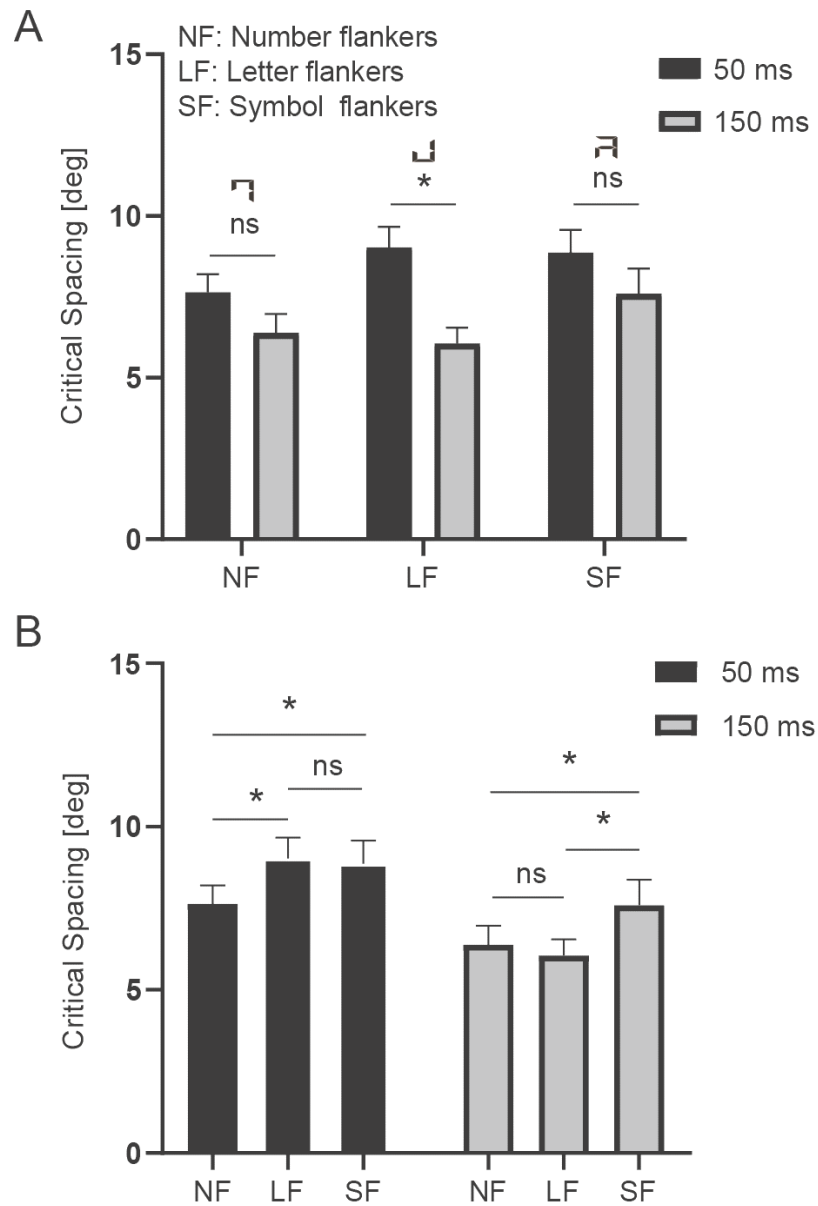


Fig. 2.3. Results for Experiment 1. Number flanker is abbreviated as NF, letter flanker is abbreviated as LF, Symbol flanker is abbreviated as SF. Error bars represent SEM. Black - 50 ms exposure time, grey - 150 ms exposure time. The symbol * means there was a significant difference, $p < 0.05$, ns means no significant difference in this pair

2.4 Experiment 2 The temporal property of flanker category

In experiment 2, we introduce the concept of interstimulus interval (ISI). ISI is the time between stimulus presentations. Here, ISI is used to describe the interval between the target stimulus and masking. In this experiment, ISI is used to study the occurrence time of flanker category effect. To compare the results between two exposure time conditions, SOA was defined as a time parameter which equals the sum of ISI and exposure time.

2.4.1 Method

Experiment 2 consisted of three conditions, included two exposure time (50 ms and 90 ms), three levels in flanker category (number, letter, and symbol flanker), and five levels in the interstimulus interval(0, 30, 60, 90, and 200 ms), totally 30 levels. Each level consisted of 75 trials, which were arranged randomly. Experiment was divided into fifteen blocks. Each block included 150 trials in all experiments.

Experiments 2 used the same procedures as Experiment 1, except for a 50ms masking after target stimuli, and an interstimulus interval for 0-200ms displayed between target stimuli and masking stimuli. In additional, the spacing between the target and the flanker is 5 deg in experiment 2. The experiment was repeated three times, and the data were averaged. The procedure of experiments was presented in Fig. 2.4.

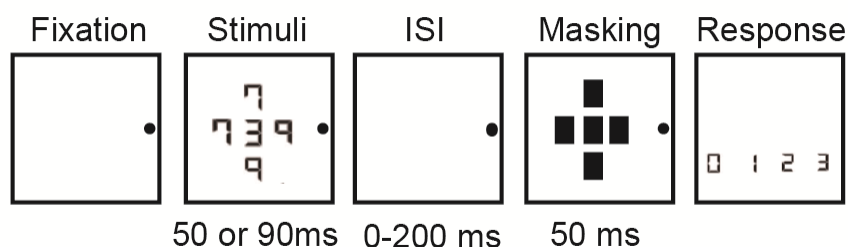


Fig. 2.4. the schematic of experimental flow in experiment 2. Several target surround by four flankers is shown randomly in the right or left on the screen. ISI was changed in different experiments. The flanker type conditions are the same in all experiments, including number flanker, letter flanker, and symbol flanker.

2.4.2 Results and discussion

Accuracy

The performance of the task was calculated by accuracy. The results are presented in Fig. 2.5A and 2.5C. A repeated measure analysis of variance (ANOVA) uncovered a significant effect of exposure time [$F(1, 5) = 103.924$ $P < .001$] and ISI [$F(1.563, 7.813) = 8.915$ $P < .001$]. There is a marginally significant of category [$F(1.588, 7.813) = 7.941$ $P = .107$]. The interaction test was significant between exposure time and ISI [$F(2.039, 10.197) = 6.348$ $P = .016$]. However, no significant differences were found in other conditions. Greenhouse–Geisser corrections were applied.

Category index

To describe the intensity of the flanker category effect, we introduce the concept of category index, which refers to the change of accuracy from different categories of flankers.

$$\text{Category index} = \text{Acc}_{\text{letter}} - \text{Acc}_{\text{number}}$$

The results of category index are displayed in Fig 2.5B and 2.5D. SOA is the sum of exposure time and interval time. The peak of category index was found when SOA about 145ms (average of 140ms and 150ms).

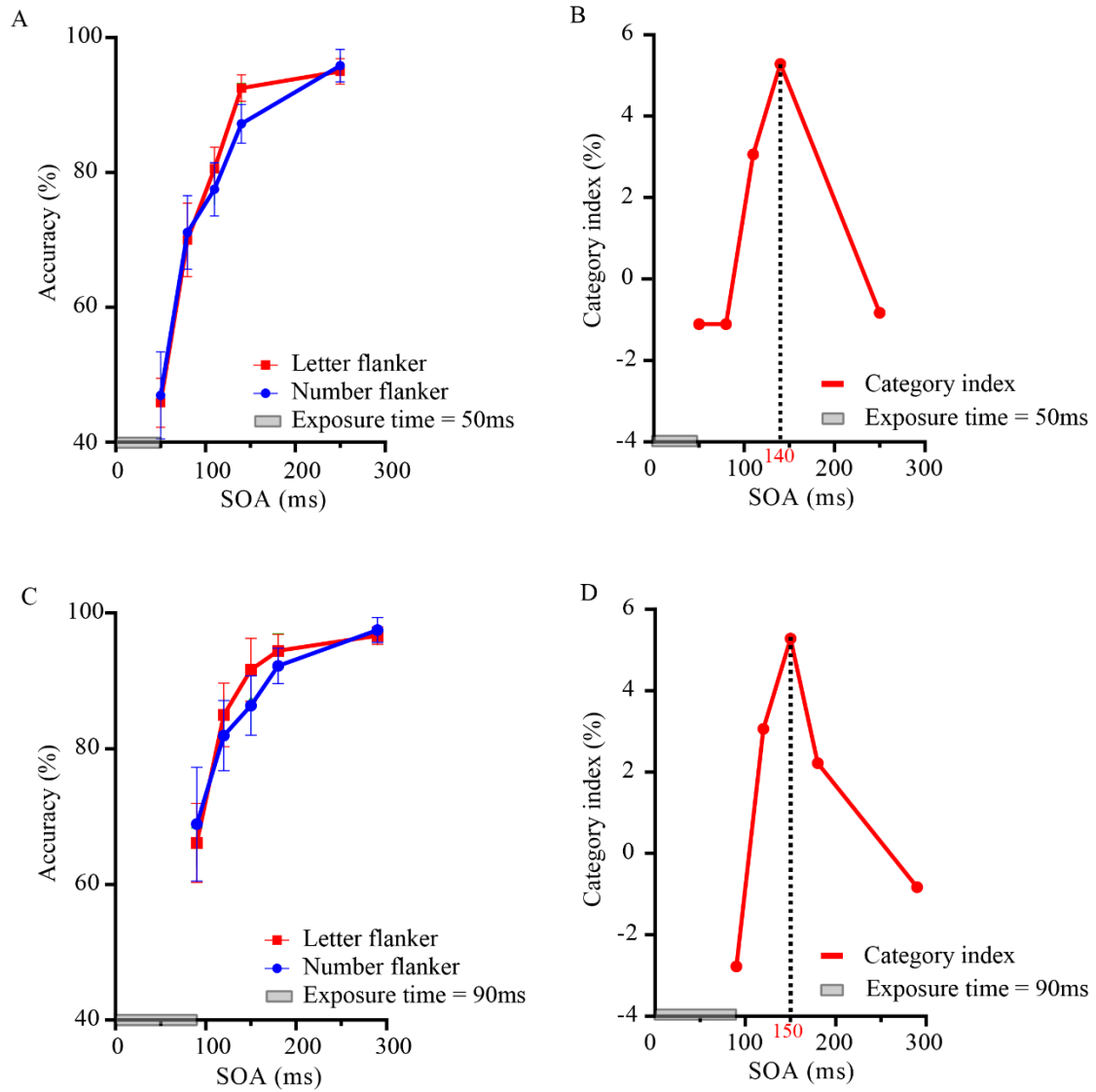


Fig. 2.5. The accuracy vs SOA function and category index vs SOA function. A for the accuracy function for the letter flanker and number flanker condition (exposure time 50ms). C for the category function which indicates the difference between the two accuracy functions (exposure time 50ms). B for the accuracy function for the letter flanker and number flanker condition (exposure time 90ms). D for the category function which indicates the difference between the two accuracy functions (exposure time 90ms).

2.5 Discussion

As mentioned in the introduction, processing time might be an important factor influencing multi-level crowding. One of the aims of this study was to explore the effect of high-level information on crowding intensity at different processing times. The core hypothesis is that the effect of flanker category might be modulated by processing time. To test our hypotheses, we designed four experiments.

In experiment 1, we explored the temporal property of flanker category effect to replicate and extend category effect documented by a previous study [6]. The experimental data suggested that flankers' category information cannot reduce crowding when the exposure duration is short. This result may be explained by the fact that the letter flankers were recognized as meaningless clutter. However, the number flankers are easily integrated into one target with a number target. Therefore, the number flanker's crowding is weaker than letter and symbol flanker in short exposure time. Under the condition of long exposure time, the result was consistent with the previous study.

The second question in this study was to know how the influence of category information changes with the processing depth. A series of experiments were performed. In experiment 2, we used SOA to measure category information's role and found the category effect time window.

According to the above experimental results, the most obvious finding emerged from the analysis is that the flanking category effect's peak occurs at about 145ms after the stimulus appears. This result may be explained by object recognition theory [23, 50].

CHAPTER 2 THE FLANKER CATEGORY EFFECT AT HIGH LEVEL

From the traditional views, the category effect should in the late stage of visual processing. However, from the perspective of occurrence time, flanker category effect starts from the grouping stage. The theory of feedforward and feedback processing might explain this visual phenomenon more clearly [51]. Flanking category effect occurs on mid-level vision, and it adopts a short time window to pick out separate objects [52].

Another important finding is that masking stimulus determines the temporal property of flanker category effect. This result is in accord with the previous research, which shows that masking method is a precise paradigm to explore the temporal question [31].

These findings suggest that multi-level crowding is a dynamic process affected by processing time. These are particularly useful results which might help others to design an experiment of temporal studies using imaging and electrophysiological method.

2.6 Conclusion

The current study aimed to determine temporal factors for the flanker category effect under visual crowding. The experimental data suggest that the flanking category effect's peak occurs at about 145 ms after the stimulus appears. The masking stimulus also determines the temporal property of flanker category effect. This study's contribution has been to confirm that processing depth can adjust the influence of high-level information on crowding. The strengths of the study included an in-depth analysis of category effect under various temporal conditions. Although the study has successfully demonstrated that the category effect is modulated by processing time, it has certain limitations in exploring brain mechanisms. Further electrophysiology studies need to be carried out to validate the temporal property of multi-level crowding.

Chapter 3 A comparison of the mechanisms underlying the memory color violation and the spatial configuration violation

Summary

We live in a world where most objects appear in specific contexts which allow for context-driven predictions or expectations. When the positions of objects in the scene are disrupted and placed in a cluttered form, the recognition speed and accuracy of target objects are reduced. In order to explore the mechanism underlying the violation effect, we investigated the influence of memory color, and 3d depth scene using the ERP and sLORETA method.

The significant difference between valid and invalid stimuli was observed at 200 and 450ms in color condition and was observed at 200ms to 400ms in depth condition. The significant difference between valid and invalid stimuli were found in color (425-480ms, Sub-Gyral) and depth condition (180-250ms, IPL; 425-480ms MFG) by using sLORETA analysis. The significant difference between valid and invalid stimuli was mediate at theta and beta band in color condition, and alpha and beta band in depth condition by using time frequency analysis. The significant difference from sLORETA analysis in cross spectrum is mediate at inferior frontal gyrus (IFG) in theta band (4–7 Hz) and superior parietal lobule (SPL) in beta band (16.5–20 Hz) under color condition.

According to the results of ERP and sLORETA. We propose that the violation effect

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

occurred in the different time window and different brain areas among memory color and 3d depth scene, suggesting that the violation of color memory and 3d depth scene are mediated by different brain mechanism.

3.1 Background

We live in a world where most objects appear in specific contexts which allow for context-driven predictions or expectations. Using prior experiences and knowledge about the specific contexts, we can detect or recognize more accurately and more rapidly thousands of objects in a cluttered scene, despite the variability of positions or changes in object occlusion.

When the positions of objects in the scene are disrupted and placed in a cluttered form, the recognition speed and accuracy of target objects are reduced. As the number of violations in a scene was increased, subjects' accuracy in detecting target objects in the scene decreased, as did the time required for subjects to judge that there was something wrong with the scene [53-55].

Object search and recognition are facilitated when the object is in the expected context, position and size. [56-61].

On the other hand, when objects are seen in an incongruent context or with unfamiliar distracter objects, visual performance is reduced because prediction or expectation is hindered [62-71].

As for the neural correlates underlying the context-driven prediction or expectation, a lot of knowledge have been accumulated by many imaging studies (fMRI and MEG).

It has been shown that prediction or expectation are contributed by the primary visual cortex [72, 73], sensory cortices [74-78], hippocampus [72, 79], parahippocampal and retrosplenial [56, 80-84].

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

A lot of previous studies are expected to converge towards the understanding of context driven prediction or expectation processing. By using different stimulus configurations and tasks to explore visual processing models and analyze the properties of visual performance in valid or invalid contexts. Thus, the processing mechanisms for context-driven prediction or expectation processing are investigated.

Indeed, as for the time course of context-driven prediction or expectation effect to visual performance on which we focus in the present study, there are only few studies [59, 85, 86].

Further, little is investigated whether gamma-band activity in the context driven prediction or expectation processing may serve as a metric to unravel the mechanism underlying context driven prediction or expectation processing. Here, we provide an additional test of the visual cue effect on visual object categorization by recording EEG while subjects are judging the contextual validity of the visual contexts and its analysis of gamma-band activity. We used a sequential context-object design for two different contexts of color and 3D structure of objects. Our sequential context-object design was based on our previous fMRI study which has shown that connectivity in cortical pathways [87], and the dynamic activation patterns of these pathways that depend on visual context.

We asked the following questions: How does context congruency influence ERPs wave and gamma-band activity in a visual categorization task? How are differences in the context effect among two contexts of color, and 3D structure of objects? What is the underlying neural mechanism for each of the two contexts?

3.2 Method

3.2.1 Participants

A total of 20 subjects (1 female) from Okayama University participated in the experiments. 20 subjects took part in the color and depth experiment (19 males, mean age 22.6 (SD=1.8) years; all 15 were right-handed). 19 participants were right-handed, and 1 was left-handed. All participants had normal hearing, had normal or corrected-to-normal vision, and normal trichromatic color vision according to the Farnsworth Munsell 100 hue tests, and reported no history of neurological or psychiatric illness. All participants signed the style (research) No. 1-1 consent form of Okayama University and completed the experiment upon their approval of the experimental content.

3.2.2 Stimuli

We focused on the color and 3D depth conditions, using stimuli with valid and invalid backgrounds. A set of visual stimuli were used in this experiment which was applied to a previous fMRI study. The visual target pictures at a visual field of $8^\circ \times 8^\circ$ (color condition), $15^\circ \times 15^\circ$ (depth condition) in a gray background. To find the modulatory effect, we used a pair of contrast conditions for two kinds of context: natural vs. unnatural color and normal vs. abnormal 3D scene. For each set of objects, 20 valid or invalid stimuli were presented.

The stimulus was applied to a previous fMRI study, where contextual validity of stimuli was evaluated. A series of the pre cue stimuli were created according to the

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

requirements of experimental design. The grayscale pictures consistent with the semantic of targets were used as the pre cue stimuli under color conditions. The black spots replaced human symbols as the pre cue stimuli under depth conditions.










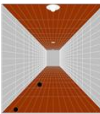








Condition	Examples of pre cue (incomplete information)		Examples of Target				
Color condition	Fruit and vegetable without color		Valid fruit and vegetable				
			Invalid fruit and vegetable				
3D depth condition	Depth sense without human like shape		Valid depth sense				
			Invalid depth sense				

Fig. 3.1. shows the stimulus parameters of the present experimental stimuli. For color condition, the visual target pictures that were embedded within the frame with 8° (visual angle) vertical and 8° horizontal were presented in the gray background. For depth condition, the visual target pictures that were embedded within the frame with 15° (visual angle) vertical and 15° horizontal were presented in the gray background. Valid: familiar, natural, plausible and consistent. Invalid: unfamiliar, unnatural, implausible and inconsistent. These are similar to those used in previous study [87].

3.2.3 Procedure

As illustrated in Fig. 3.2, after the presentation of a fixation icon (500 ms), the prime was presented for 200 ms, and the target was presented for 200 ms successively with

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

an interstimulus interval (ISI) of 2000 ms. A white fixation was presented for 600-2100 ms to collect EEG signals. And then a black fixation was presented for 500ms. Participants were asked to determine whether the target was naturally or unnaturally matched after the black fixation appear.

Experimental flow

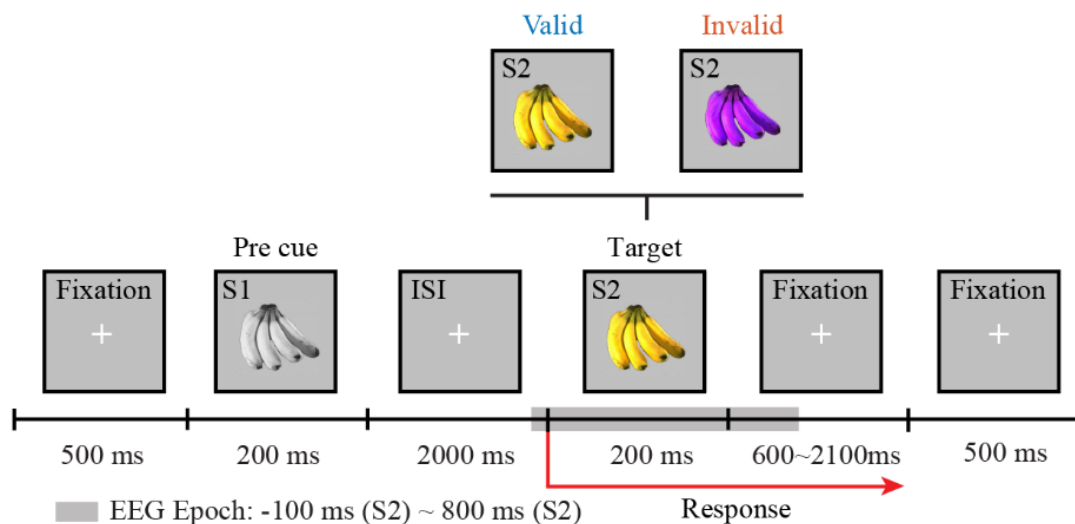


Fig. 3.2. A pre cue (S1) and visual target (S2) were presented for 200ms, respectively, following the presentation of a fixation icon for 500 ms. The interstimulus interval (ISI) was 2000 ms. Participants were asked to respond as fast and accurately as possible after visual target was displayed. Participants were required to visual target is valid or invalid. Participants' ERPs were recorded during the task.

3.2.4 EEG recordings

Continuous EEG data were collected from 32 scalp sites using sintered Ag/AgCl electrodes mounted on an elastic cap. Electrodes were positioned according to the

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

international 10-20 standard (Jasper, 1958). All electrodes were referenced to linked earlobes, and the ground electrode was positioned on the forehead. Two electrodes were positioned below and on the sides of the eyes to record the EEG data. Skin-electrode contact impedance levels were maintained below 5 k Ω . EEG signals were recorded continuously at a sampling rate of 500 Hz.

3.2.5 ERP analysis

The epoch of ERPs was derived starting 200 ms before target (S2) onset and lasting for 800 ms. A pre-stimulus time window of 200 ms served as baseline. Statistical analysis of the ERPs in response to picture stimuli was carried out for nine regions of interest (ROIs). Regions were defined as left frontal (LF: Fp1 and F7), middle frontal (MF: F3, Fz, and F4), right frontal (RF: Fp2 and F8), left central (LC: FC5, T7, C3, and CP5), middle central (MC: FC1, FC2, Cz, CP1, and CP2), right central (RC: FC6, T8, C4, and CP6), left posterior (LP: P3, P7, and O1), MP: Pz, POz, and Oz), and right posterior (RP: P4, P8, and O2). The obtained data were filtered in this step. The frequency of the low-pass filter was 1 Hz, and the frequency of the high-pass filter was 80 Hz. New reference was used to select the electrode to be used for analysis. Analysis was performed with reference to all the electrodes in this experiment. The obtained brain wave data were divided for the trigger point in each condition. Additionally, we set the start and end points of the division. In this analysis, the point of 200 ms before the target stimulus was presented was defined as the starting point, and the point of 800 ms after presentation was regarded as the end point. Unsuitable brain waves affected

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

by eye movements, swallowing, etc. were eliminated for each section. We removed amplitudes with a minimum of -80 μV and a maximum of 80 μV as artifacts. The brain wave data obtained for each condition were averaged in this step. The potential reference value of the EEGs was calibrated and averaged over each condition. In our experiment, the average value of the potential from -200 ms to 0 ms at the moment of target stimulus presentation was calibrated to the reference value.

3.2.6 Time-frequency analysis

The method of time-frequency analysis was based on continuous wavelet transforms (CWTs), in that EEG signals were projected from the time domain to the time-frequency domain. Time-frequency analysis was performed for each channel by convolving the data with a complex Morlet wavelet. We used the mother wave as ‘Cmor 1-1.5’ (Q increasing=1.5) for frequencies from 1 to 80 Hz (step size 1 Hz) in gamma-band activity analysis. The time-frequency analysis is based on a single-trial level and the analysis results in total power (with phase-locked and non-phase-locked signals). The baseline correction is based on the following equation: $P(t, f)_{\text{corrected}} = 100 \times \frac{(P(t, f) - P_{\text{baseline}}(f))}{P_{\text{baseline}}(f)}$. The period (-200 to 0 ms) before visual target served as baseline for all time-frequency analyses. Grand mean time-frequency results were computed over all subjects. Statistical analysis of frequency band activity was carried out for three regions of interest (ROIs). Regions were same as ERP analysis.

3.2.7 Analysis of source estimations

Analysis of source estimations used to find the difference brain activation area between congruent and incongruent conditions. We use an algorithm named ‘Standardized Low Resolution Electromagnetic Tomography’ [35]. This is a method for estimating cortical generator localization using nonparametric statistical analysis (the method corrects for multiple comparisons and does not require the Gaussianity assumption). Localization was performed in about 6000 cortical gray matter voxels of size 5 mm^3 [88].

3.3 Result

3.3.1 Results of the behavioral experiment

The behavioral results obtained from the data of all the four conditions for the 360 total trials are shown as the mean and standard deviation (SD) in Fig. 3.3. Data were analyzed using two-way repeated measures ANOVA and Holm-Sidak post hoc by using JASP 0.14 [89].

For the accuracy, the analysis for ANOVA found a significant main effect of context type ($F(1,19)=39.51$, $p < 0.001$, $\eta^2 = 0.68$) and stimuli validity ($F(1,19) = 12.58$, $p < 0.01$, $\eta^2 = 0.40$). And found a significant interaction effect [context type \times stimuli validity] ($F(1,19) = 5.63$, $p=0.03$, $\eta^2 = 0.23$). In color condition, the accuracy for the valid stimuli (mean=98%, SD=2%) were almost same ($t=0.94$, $p=0.35$) than the invalid stimuli (mean=97%, SD=3%). In depth condition, the accuracy for the valid stimuli (mean=95%, SD=4%) were greater ($t=4.21$, $p<0.001$) than those for the invalid stimuli (mean=91%, SD=6%).

For the reaction times, the analysis for ANOVA found a significant main effect of context type ($F(1,19)=37.32$, $p < 0.001$, $\eta^2 = 0.66$) and stimuli validity ($F(1,19) = 6.23$, $p < 0.05$, $\eta^2 = 0.25$). But no interaction effect [context type \times stimuli validity] ($F(1,19) = 1.01$, $p=0.33$, $\eta^2 = 0.05$). In color condition, the reaction times for the valid stimuli (mean=537 ms, SD=119 ms) were faster ($p<0.05$) than those for the invalid stimuli (mean=565 ms, SD=105 ms). In depth condition, the reaction times for the valid stimuli (mean=658 ms, SD=135 ms) were almost same ($p=0.11$) as the invalid stimuli

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

(mean=675 ms, SD=114 ms).

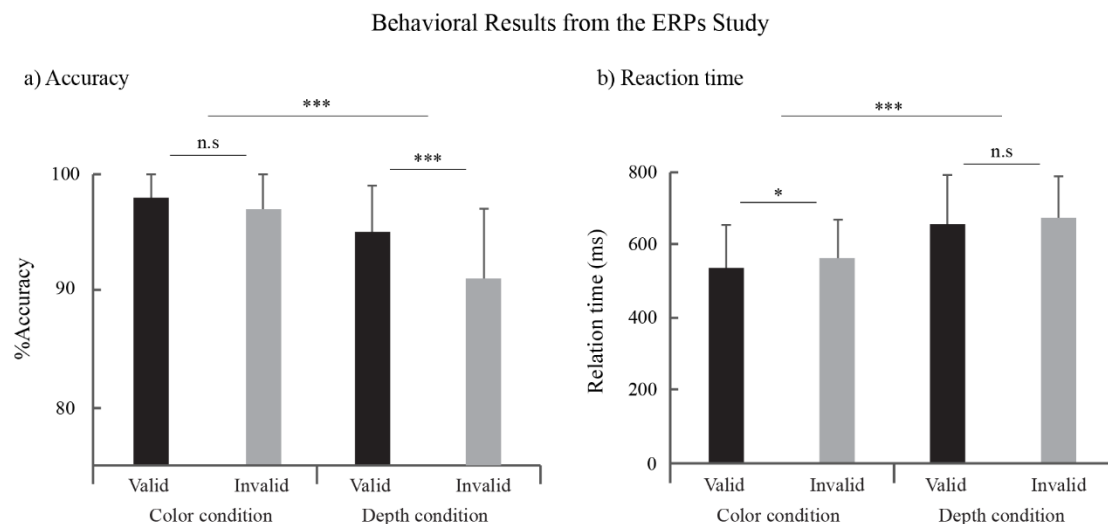


Fig. 3.3. Color condition had significantly greater accuracy and shorter reaction time than depth condition. For the accuracy, the valid condition (mean=95%, SD=4%) were greater ($t=4.21$, $p<0.001$) than those for the invalid condition (mean=91%, SD=6%) in depth condition. For the reaction times, the reaction times for the valid condition (mean=537 ms, SD=119 ms) were faster ($p<0.05$) than those for the invalid condition (mean=565 ms, SD=105 ms) in color condition.

3.3.2 Results of ERP

The grand-average ERPs time-locked to visual target across all 20 subjects are illustrated in Fig. 3.4 for color and depth condition. ERPs in response to valid (blue trace) and invalid (red trace) target stimuli are shown for the nine regions of interest. The difference in ERPs (black trace) between the valid and invalid stimuli is also shown. A cluster-based permutation test (number of permutations = 2000) was performed using the Letswave toolbox with $\alpha = 0.05$ for cluster thresholding. The black bar illustrates the time range with a significant difference. The difference reflects the difference

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

between valid and invalid stimuli clearly. A significant difference is observed about 200 ms and 400-500 ms for color condition. A significant difference is observed about 200 ms and 300-500 ms for depth condition. The time windows of significant difference are different for color and depth condition.

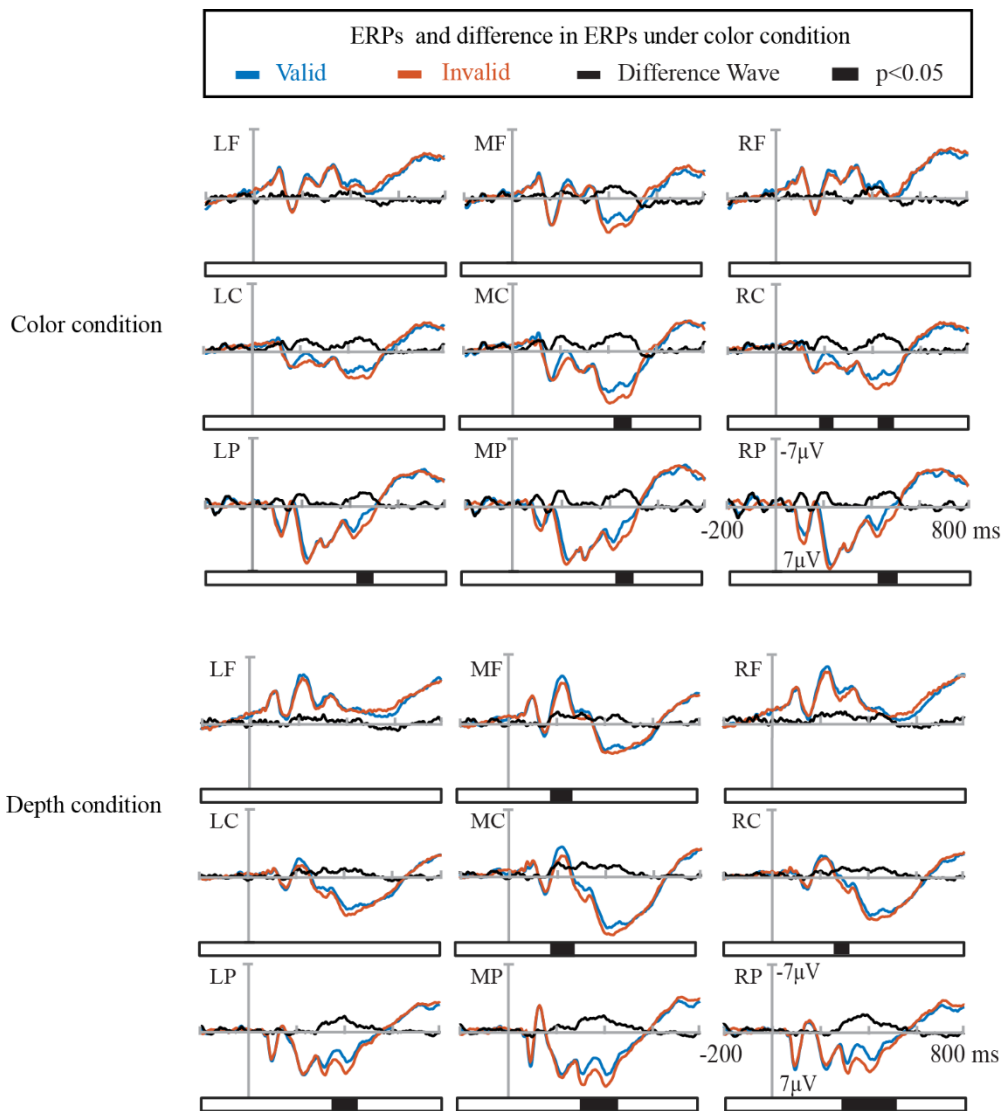


Fig. 3.4. ERPs and difference in ERPs with valid and invalid condition. ERPs in response to valid (blue trace) and invalid (red trace) target stimuli are shown for the nine regions of interest. The difference in ERPs (black trace) between the valid and invalid stimuli is also shown. A cluster-based permutation test (number of permutations = 2000) was performed using the

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

Letswave toolbox with $\alpha = 0.05$ for cluster thresholding. The black bars illustrate the time range of significant differences.

3.3.3 Results of significant difference of ERPs

Fig. 3.5 shows the significant time range and cluster t values between valid and invalid stimuli for color and depth condition. The brain topographies across three time windows have shown in right. The most pronounced difference for color condition appears in the time window of 400-550ms. But the most pronounced difference for depth condition appears in the time window of 300-400ms and 400-550ms.

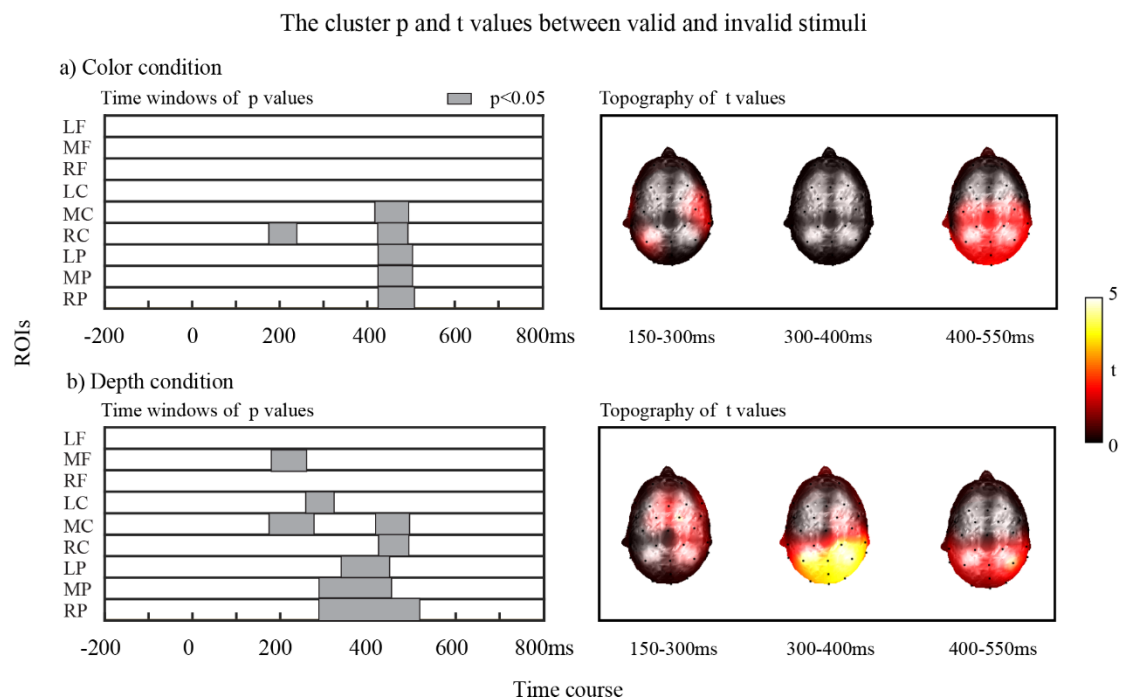


Fig. 3.5. The cluster p and t values between valid and invalid trials. The gray bar illustrates the time range of significant differences ($p < 0.05$).

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

3.3.4 Source estimations of ERPs

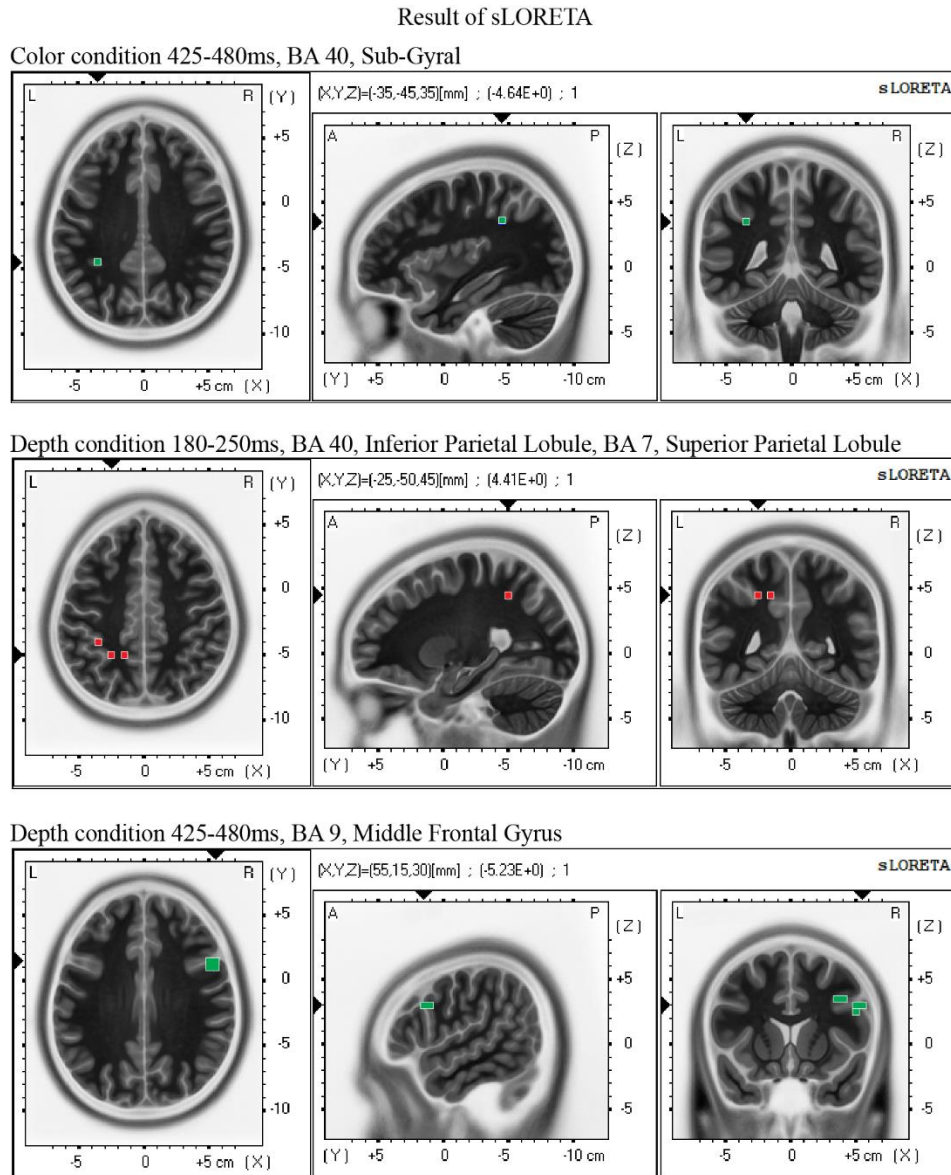


Fig. 3.6. The difference between valid and invalid condition appeared in 180-480 ms. We compared the differences in source localization between the valid and the invalid conditions. The significant difference was found in color and depth condition. Significant differences are marked in red (valid > invalid) and green (valid < invalid). Significant level is $p < 0.05$.

Visual context-related brain activities for valid and invalid stimuli were compared to examine the areas involved during the context process in color and depth condition.

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

The brain activities are displayed in Fig. 3.6. Invalid stimuli evoked larger activation in color condition, the differential activations between valid and invalid stimuli appears in sub-gyral gyri lobule of the parietal lobe (BA40) at 425-480 ms. Invalid stimuli evoked larger activation in depth condition at 180-250 ms, the differential activations between valid and invalid stimuli appears in inferior parietal lobule (IPL, BA40) and superior parietal lobule (SPL, BA7). And valid stimuli evoked larger activation in depth condition at 425-480 ms, the differential activations appear in middle frontal gyrus (MFG, BA9).

3.3.5 Result of time frequency analysis

Fig. 3.7 shows the time-frequency representation of the visual response for invalid stimuli, valid stimuli and difference at 6 ROIs (MF, LC, MC, RC, LP, MP and RP). The time-frequency representations display total power, comprising both phase-locked and nonphase-locked fractions of oscillatory activity. The black contour line in the plot in the right panel shows significant differences between valid and invalid stimuli ($p < 0.01$). In color condition, a difference in theta-band power between the valid and invalid stimuli is observed in the frequency-time windows of approximately 4-7 Hz after 400ms. A difference in beta-band power is observed in the frequency-time windows of approximately 15 Hz, 100 ms and 30 Hz, 400ms. In depth condition, a difference in alpha-band power between the valid and invalid stimuli is observed in the frequency-time windows of approximately 10 Hz 200-400ms. A difference in beta-

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

band power is observed in the frequency-time windows of approximately 18 Hz, 250 ms.

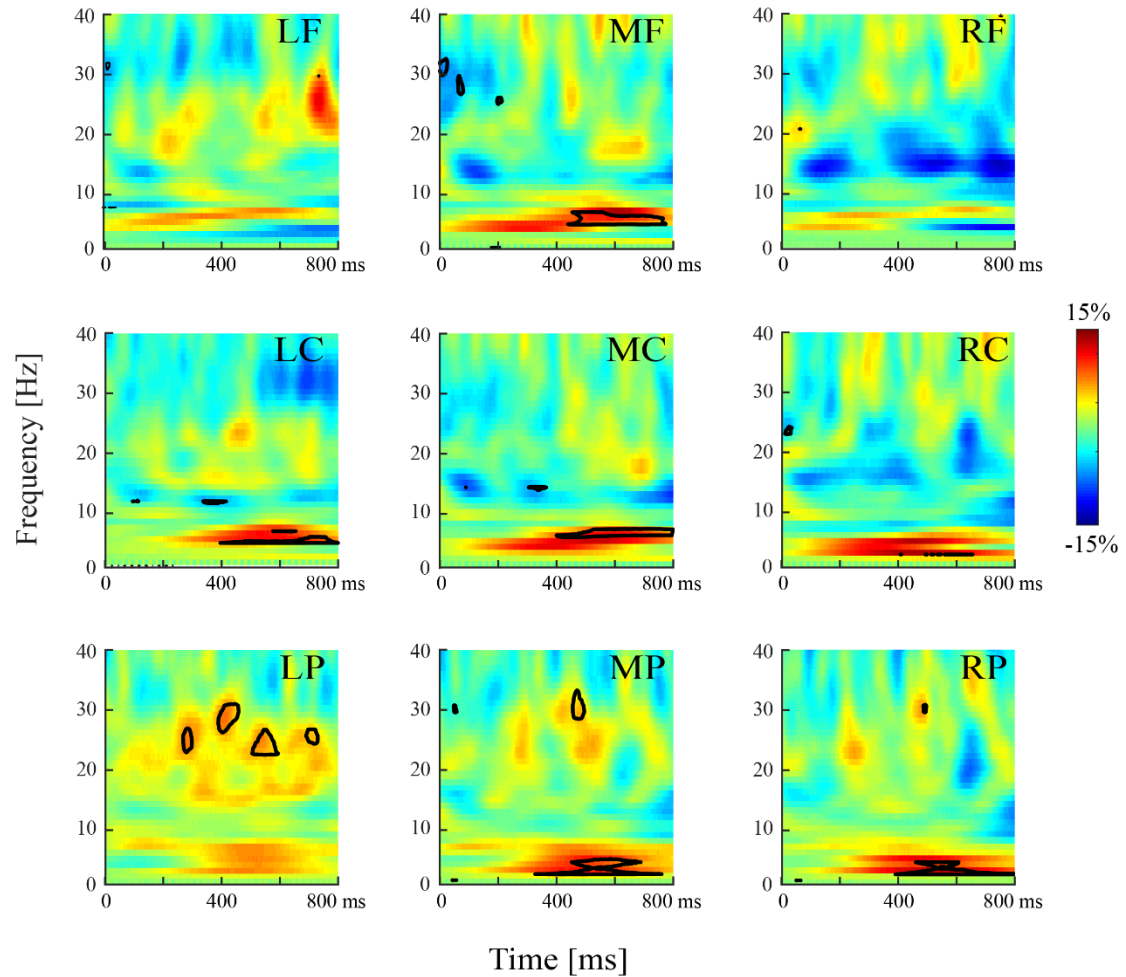


Fig. 3.7a. Time-frequency plots in color condition. The analysis was based on the difference between valid and invalid stimuli. The plots show total oscillatory activity expressed as percent change relative to baseline following the pre cue stimuli. The black contour line in the plot to the right indicates significant differences ($p < 0.01$).

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

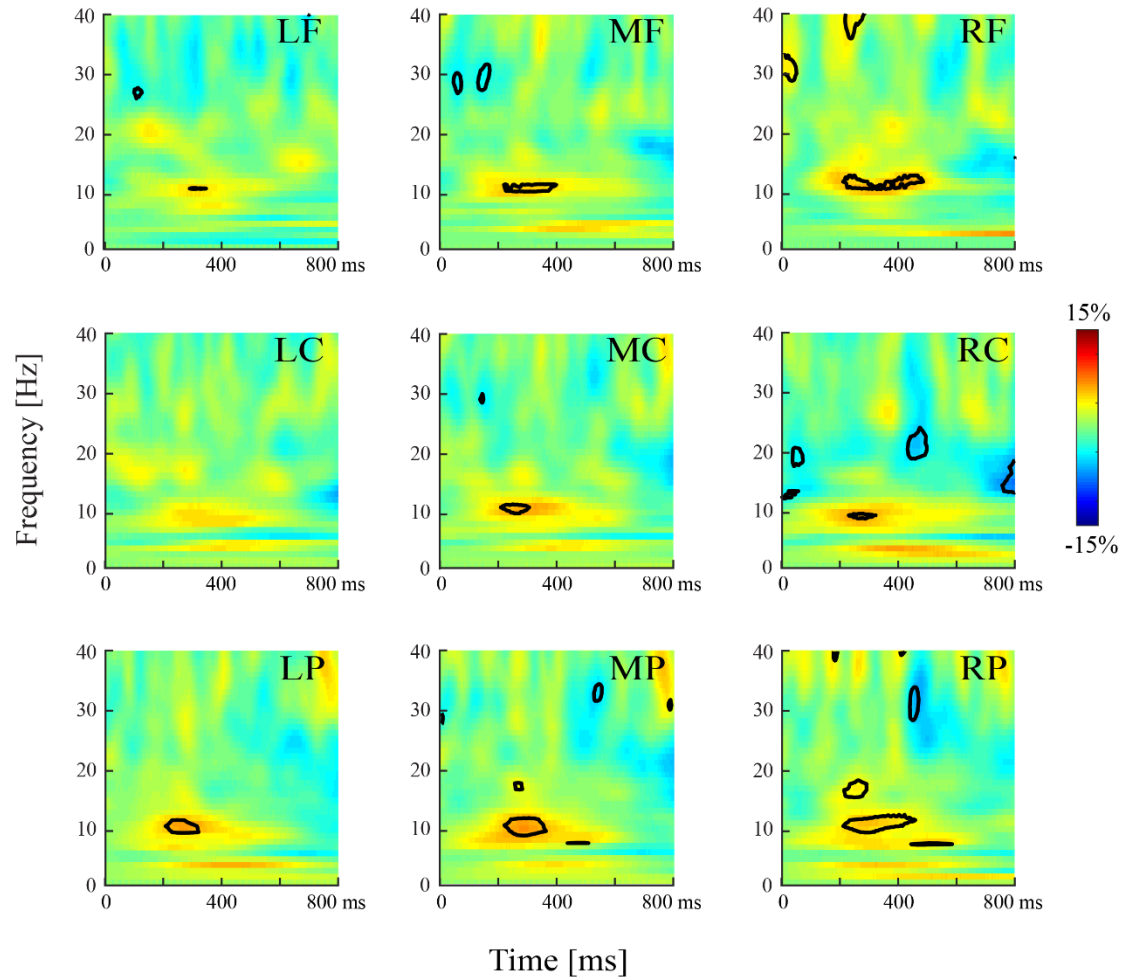


Fig. 3.7b. Time-frequency plots in depth conditions. The analysis was based on the difference between valid and invalid stimuli. The plots show total oscillatory activity expressed as percent change relative to baseline following the pre cue stimuli. The black contour line in the plot to the right indicates significant differences ($p < 0.01$).

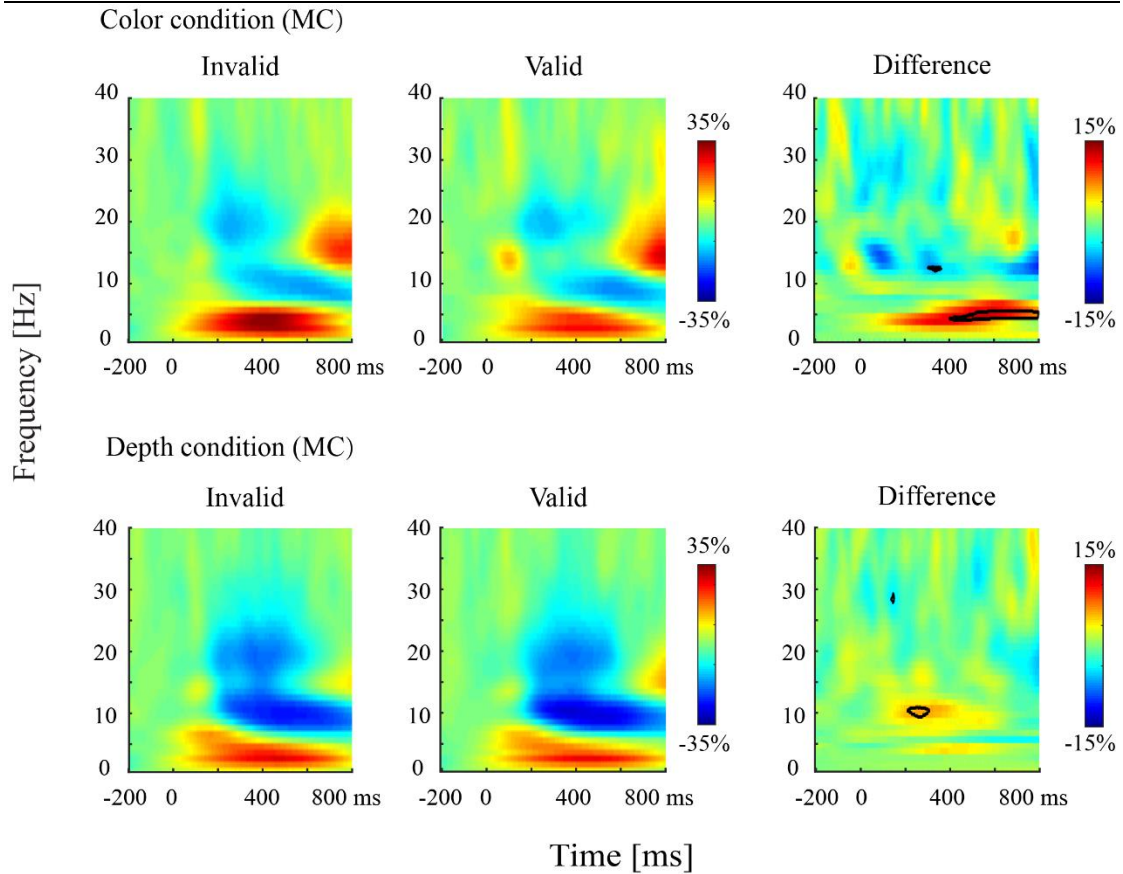


Fig. 3.7c. Time-frequency plots of valid and invalid stimuli in MC ROI under color and depth conditions. The black contour line in the plot to the right indicates significant differences ($p < 0.01$).

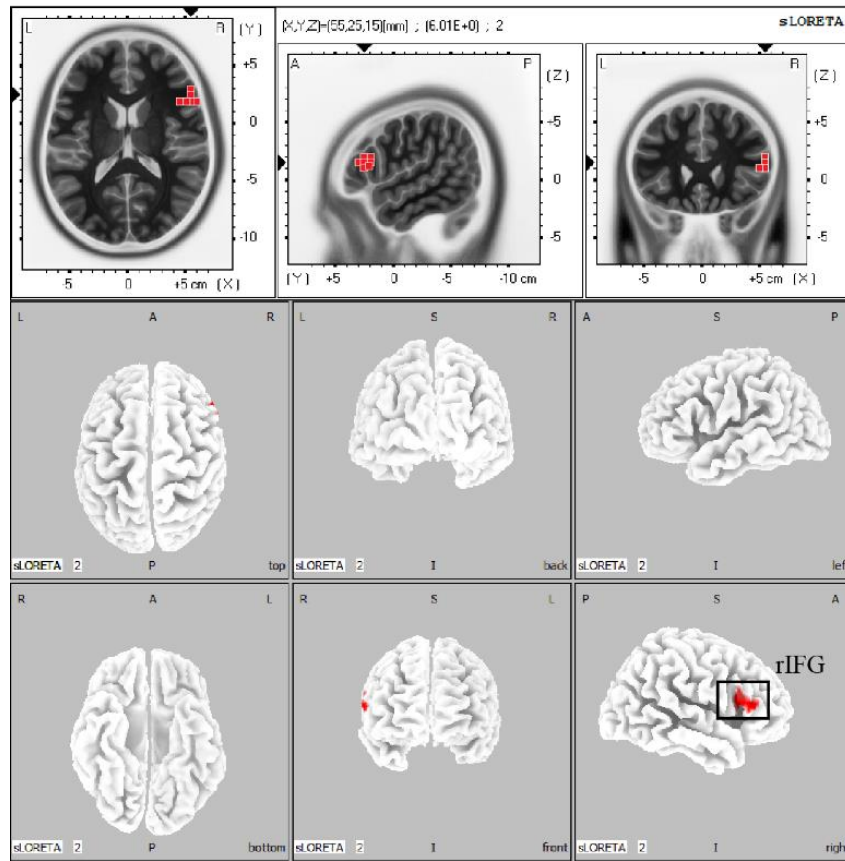
3.3.6 Source estimations of ERPs in specific band

The significant difference in specific band is observed only for color condition, sLORETA comparison demonstrated that after the valid stimuli was significant ($p < 0.05$) greater than invalid stimuli in the current source density (CSD) in inferior frontal gyrus (IFG)- BA 45 for theta band (4-7 Hz), and superior parietal lobule (SPL)- BA 7 for beta-2 band (16.5-20 Hz).

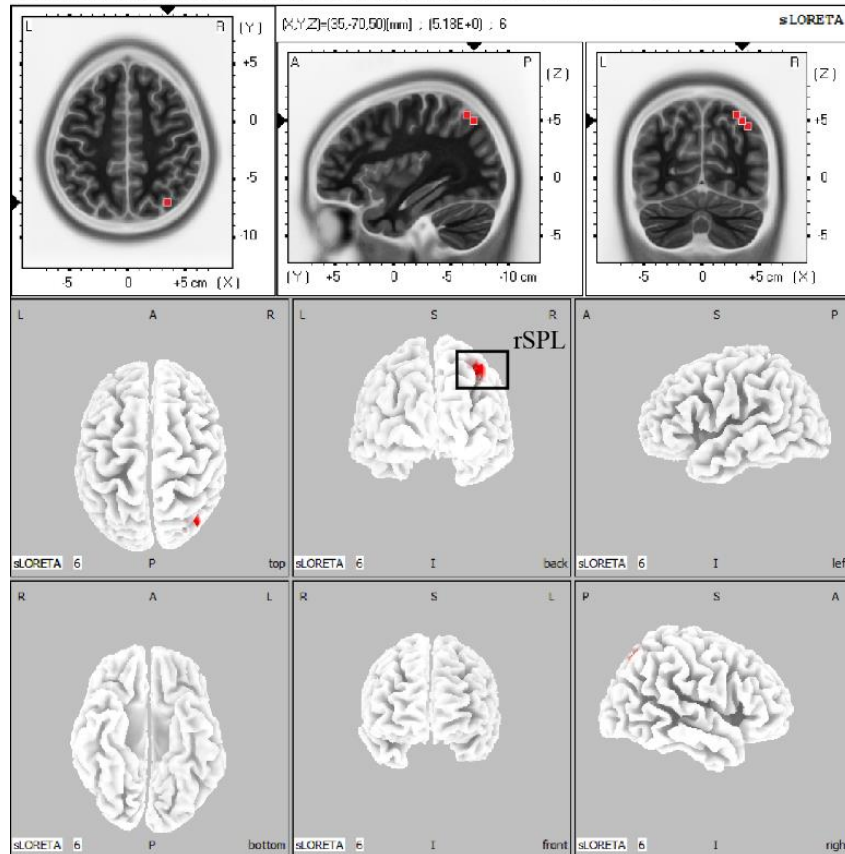
CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

Result of sLORETA in cross spectrum

Color condition Theta 4-7Hz, BA 45, Inferior Frontal Gyrus



Color condition Beta2 16.5-20Hz, BA 7, Superior Parietal Lobule



CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

Fig. 3.8. The difference between valid and invalid condition appeared theta and beta band.

We compared the differences in source localization between the valid and the invalid conditions.

The significant differences were found in color condition (theta band rLFG and beta band rSPL).

Significant differences are marked in red (valid > invalid) and green (valid < invalid).

Significant level is $p < 0.05$.

3.4 Discussion

3.4.1 Comparison to behavior data indicate different context processing

We found color condition had significantly greater accuracy and shorter reaction time than depth condition. There was a significant difference of stimuli validity in accuracy for depth condition, and a significant difference in reaction time for color condition. This finding is in line with previous reports showing that different visual context affects the performance of visual object recognition. Our behavioral data corroborate the findings of a great deal of the previous work of contextual modulations of object-level information on the behavioral study. Our results enhance the current knowledge of the visual context of color and depth.

About color studies, researchers found the visual identity of an object has a measurable effect on color perception, that suggesting there is a robust two-way influence between memory colors and features of visual object[90]. The results of reaction time in this study show the Promoting effect of stimulus consistency on memory color judgment.

About depth studies, researchers found modulations of behavioral depth judgments based on object plausibility (valid or invalid stimuli) [91], that suggesting the object recognition plays an important role in encoding of 3D-information. The accuracy results in this study also proved that stimuli validity can improve the preference of 3D-depth processing.

Direct comparison studies about color and depth plausibility were sparse, but a recent

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

study compared the difference of two visual context. We found reaction time in color condition was significantly shorter than those in depth condition [87]. Here consistent with the previous studies, Wu et al also found that the performance difference of behavior judgment between color and depth condition [87]. We found color condition had significantly greater accuracy than depth condition. However, this result has not previously been described in Wu's results. A possible explanation for this is that we used pre cue in validity judgment task. Compared with position information, contour information triggers a stronger priming effect.

3.4.2 Comparison to validity-related ERP data indicate different context processing

As shown in fig 5, we found significant differences in ERPs between valid and invalid stimuli in some ROIs and time windows. Note here ERPs for the invalid stimuli are larger in amplitude than those for the valid stimuli, indicating that the unnatural (unfamiliar, implausible, low-probable) stimuli induce larger signals than the natural (familiar, plausible, high-probable) stimuli. This nature of trend which was found first by Sutton, Braren, Zubin and John [92], has been reported in many research fields of vision, audition and language by a body of previous studies (for a review) [93, 94]. For example, 33 % sound and light induce larger amplitude in ERPs than 66% sound and light [92]. Inverted face, impossible human pose and blue colored face induce larger amplitude in ERPs than upright face, possible human pose and natural colored face, respectively [95]. Deviant words induce larger amplitude than normal words [96]. A model for explaining the nature of trend was proposed [97]. However, the underlying

mechanism remains unresolved. The neural correlates about the present finding shown in Fig. 3.5 are addressed in the next session.

3.4.3 Comparison to validity-related ERSP data indicate different context processing

Prior studies have noted the importance of event-related spectral perturbation (ERSP) in the human cognitive study. One of the aims of this study was to compare neural mechanisms of stimuli validity of color and depth condition. We found a significant difference in theta-band power under color condition and a significant difference in alpha-band power under depth condition. The differences were also noted in beta frequency band under both color and depth band.

In color condition, our results show that the invalid stimuli evoked stronger brain oscillations in theta frequency band than those evoked by valid stimuli under multiple ROIs. The results of theta-band power for color condition are consistent with Shin who demonstrated the relation between P300 and event-related theta-band [98]. However, it has been suggested that P300-theta relation should be treated with caution unless the single-trial-based method was used [99]. The possible interference of methodology was ruled out because we performed a single-trial-based analysis. Another possible explanation for this is that the relationship between working memory and theta band activity. Invalid stimuli conflicts with memory representations induced by pre cue. Our results show that the invalid stimuli evoked stronger brain oscillations in theta frequency band than those evoked by valid stimuli under multiple ROIs. Compared to decision on valid stimuli, decision on invalid stimuli require more memory search. This

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

seems to predict that the key time window for color validity decision is after 400 ms and may be driven by working memory information.

In high-level cognitive studies [100], the beta band is related with decision making. The increase in beta band oscillations predicts stronger decision processing. The decision on invalid seems to require more processing than the valid condition.

In depth condition, a significant difference in energy oscillation was observed at alpha and beta bands in the validity treatment under depth condition. About alpha band oscillation, three aspects are interesting to highlight from a functional point of view.

Firstly, it is well known that the reduction of the alpha band is highly related to the focusing of attention. (Reference list) ERS (event-related synchronization) of the alpha band means inhibition and ERD (event-related desynchronization) means release inhibition [101, 102]. Our results show that the valid stimuli evoked weaker brain oscillations in alpha frequency band than those evoked by invalid stimuli under multiple ROIs. This result suggests that the recognition of valid stimuli seems to require more attention. In this experiment, visual spatial attention modulated the scene perception of subjects. Two factors are determining whether a scene stimulus is valid or invalid: the position of the human-like symbols and the size difference of the human-like symbols. Both valid and invalid stimuli are prompted by the cue stimulus with position information of the human-like symbols. So, the symbol position in the scene does not cause attentional differences between valid and invalid stimuli. We suggest that the alpha oscillation difference may originate from the size difference of the human-like symbols.

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

Secondly, our findings are consistent with the hypothesis that the increase in alpha activity is associated with functional inhibition. That is, ERS reflects inhibition and ERD the release from inhibition [101, 102]. From the results, both invalid and valid stimuli show the ERD in the alpha band, which implies a release of functional inhibition. However, valid stimuli showed the weaker alpha oscillation, which may imply stronger functional activation. The reduced alpha band oscillations seem to predict the validity of the 3D depth visual context.

Lastly, our findings may be related to the knowledge system. The knowledge system is a storage system of knowledge that contains multiple procedural and declarative information, and the knowledge system is related to both working memory and long term memory [102]. The researchers found that during long term memory retrieval, higher semantic integration of information implies stronger ERD [103]. Our results suggest that valid stimuli perform more long term memory extraction from the knowledge system. The result predicts that processing of 3D depth visual context in valid-stimulus decisions is associated with the spatial rules from long-term memory.

In additional, the results in the beta frequency band (20-30 Hz) showed that the valid stimuli evoked stronger oscillations than the invalid stimulus. This result supports that the valid stimulus requires more decision processing in the depth condition.

In summary, we suggest that subjects use working memory information to measure the plausibility of invalid stimulus in color condition. Subjects expend more energy in the decision-making processing of invalid stimuli. In contrast, the decision of the valid stimulus is relatively simple. In depth condition, subjects tend to pour more attention to

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

valid stimuli by using long-term memory. Subjects expend more energy in the decision-making process of valid stimuli. The results alpha-band power for depth condition can be explained using the relativity of long-term memory and alpha band activity.

3.4.4 The relativity of brain region in different context processing

This part focused on source localization of P300 in color and depth context processing. We expand findings by Wu using fMRI method [87]. Invalid stimuli triggers stronger P300 amplitudes.

For color condition, the P300 generators were found in BA40 in 425-480ms. Consistent with the results of other researchers who located the P300 in parieto-occipital areas [104]. The subject displayed stronger allocation of attention to invalid stimuli. Further analysis found that a significant difference is mediate at right inferior frontal gyrus (rIFG) in theta band (4–7 Hz) in color condition. This result may be explained by the fact that the theta-related processes in the right inferior frontal gyrus (rIFG, BA45) during response inhibition[105]. Working memory are strongly dependent on theta band activity and can affect inhibitory control processes [106-108]. In additional, a significant difference is mediate at right superior parietal lobule (rSPL) in beta2 band (16.5–20 Hz) in color condition. The superior parietal lobule is involved in mental imagery and recall of personal experiences [109]. These results may be due to the recollection of memory color for valid stimuli.

For depth condition, the N2 generators were found in inferior parietal lobule (IPL, BA7) and superior parietal lobule (SPL, BA40) in 180-250ms. The P300 generators

CHAPTER 3 THE COMPARISON BETWEEN MEMORY COLOR AND 3D CONFIGURATION

were found in middle frontal gyrus (MFG, BA9) in 425-480ms. The MFG have been found to be involved in the generation of the P300 [110].

3.5 Conclusion

The significant difference between valid and invalid stimuli was observed at 450ms in color condition and was observed at 200ms and 400ms in depth condition. The significant difference between valid and invalid stimuli was mediate at theta and beta band in color condition, and alpha and beta band in depth condition by using time frequency analysis. The significant difference between valid and invalid stimuli was found in color (425-480ms, sub-gyral gyri lobule of the parietal lobe, BA40) and depth condition (180-250ms, IPL, BA40, SPL, BA7; 425-480ms MFG, BA9) by using sLORETA analysis. The significant difference between valid and invalid stimuli was mediate at theta band in color condition, beta band in depth condition) by using time frequency analysis. The significant difference is mediate at right inferior frontal gyrus (rIFG) in theta band (4–7 Hz) and right superior parietal lobule (rSPL) in beta2 band (16.5–20 Hz) in color condition. We propose that the violation effect occurred in the different time window and different brain areas among memory color and 3d depth scene, suggesting that the violation of color memory and 3d depth scene are mediated by different brain mechanism.

Chapter 4 The mechanism underlying the interaction processing between audition and vision by violation method

Summary

Semantic congruency has a facilitatory effect on multisensory integration. However, the neural mechanisms underlying auditory to visual interaction in cross-model priming have been unexplored. In particular, the role of implicit information (semantic size) in cross-model integration is still poorly understood.

We used the EEG method and an auditory-to-visual priming paradigm in this study. The spatial and temporal dynamics of semantic congruency between auditory-visual natural object stimulus pairs were evaluated using time-frequency analysis (TFA) and standardized low-resolution electromagnetic tomography (sLORETA).

The event-related potentials (ERPs) between 250 ms and 350 ms for the incongruent condition showed a negative-going deflection compared to those for the congruent condition, the significant difference of ERPs is mediate at the frontal lobes and the occipital lobe, and the significant difference of lateralization is mediate at left middle frontal gyrus (IMFG) and right superior frontal gyrus (rSFG). The congruency effect in gamma-band activity was observed in the frequency-time window of 60-70 Hz at 200 ms. Here the result of ERPs classified by prime-target pairs was analyzed further. The N400 effect was depended on the semantic size of prime-target pairs. The dependency

CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

suggests the congruency effects determined by the semantic size difference in auditory to visual priming paradigm.

These findings could be accounted for by a hypothetical model in which the semantic information driven by the auditory prime and the information of the visual target may be integrated during visual object processing.

Keywords: object categorization, auditory-visual priming; ERP recording; auditory-visual interaction; gamma-band activity; top-down and bottom-up interaction, dependency on prime-target type.

4.1 Background

We live in the natural environment rich in a wide variety of information. The information does not necessarily present simultaneously. For example, when we see lightning at a distance, we anticipate thunder before hearing it. We then hear it faster than we would have without seeing the lighting. Our auditory system is facilitated by anticipation. When we hear flapping wings, we find birds more easily and faster. Our visual system is facilitated. To make sense of the real world, our brain must process information through different sensory inputs and integrate them by using expectation or prediction processing across vision, audition and somatosensory(touch, smell, and taste) systems. One approach to studying expectation processing across different modalities is to examine how the modalities interact in object recognition by using a crossmodal prime paradigm. In experimental setups, such expectation processing is often investigated in terms of crossmodal association effects such as a priming effect on a subject's response to a target stimulus with different modalities. For instance, how does a semantic stimulus influence the processing of a visual stimulus [111]? How does a visual or auditory prime influence the processing of an auditory or visual stimulus [112]?

Regarding the visual-auditory association effects that we are particularly interested in here, many previous studies have shown that visual (or auditory) primes influence the perception of auditory (or visual) objects in a variety of stimuli (auditory stimuli such as naturalistic sounds, spoken words, environmental sounds or object specific sounds and visual stimuli such as printed words, line drawings, pictures of natural

CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

objects and displayed objects) and tasks such as detection, categorization, identification and recognition of targets and matching between primes and targets.

Many previous studies [113-118] are expected to converge toward the understanding of expectation processing across visual auditory modalities. There remains the possibility that accumulating structural properties of expectation processing further by using a wider range of stimulus conditions and different tasks and by exploring a possible model may open a new perspective and lead to deep insight into the object encoding mechanism.

Indeed, the time course of vision-to-audition crossmodal processing in which visual prime stimuli facilitate auditory object recognition has been well investigated in by many ERP studies [119-127], and an underlying model was proposed [128].

However, in little information is known about the time course of the auditory priming effect on visual recognition by ERP or MEG method [84, 124, 129-133].

Furthermore, few studies have investigated whether gamma-band activity in the context of audition-vision integration may serve as a mechanism for the integration of object features across audition and vision. Here, we provide an additional test of the auditory prime effect on visual object categorization by recording EEG and its analysis of gamma-band activity. We asked the following questions: How do auditory-to-visual congruency influence ERP waves and gamma-band activity in a visual categorization task, and what is the difference between the audio priming effect on visual categorization and the visual priming effect on auditory categorization? How are the congruency effects determined?

CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

As a consequence, we found the differences in ERP between the congruent and incongruent conditions observed in the time window 250 and 350 ms after target stimulus onset and the differences in the early gamma-band activities in the frequency-time window of 60-70 Hz-200 ms. in the middle frontal (MF), middle central (MC) and middle posterior (MP) regions. In addition, the N400 effect is dependent on the semantic size difference between the auditory prime and visual target. The dependency could be accounted for by the interaction between the visual target information and the semantic visual information derived by the auditory prime.

4.2 Methods

4.2.1 Participants

A total of 36 subjects (5 female) from Okayama University participated in the experiments. All subjects participated in the evaluation test of stimuli. Fifteen male students participated in the ERP study, which is a usual size for an ERP study. Fourteen participants were right-handed, and 1 was left-handed. All participants were native speakers of Japanese, recruited from the Department of Engineering, recruited and tested during the academic years of 2018 to 2019. Their mean age was 22.6 years, with a range 21~25 years. All participants had normal hearing, had normal or corrected-to-normal vision and reported no history of neurological or psychiatric illness. We obtained ethics approval from Okayama University. All participants signed the style (research) No. 1-1 consent form of Okayama University and completed the experiment upon their approval of the experimental content. One participant was excluded from the data analysis due to artifacts in the EEG recording, and one was excluded from the behavioral data analysis because of the missing reaction time.







4.2.2 Stimuli

A set of auditory and visual stimuli employed in the present ERP experiments were selected by four tests of confident auditory categorization, identification, and correct auditory-visual matching of natural objects (animals). The sounds of six animals served as primes (S1s), and visual pictures of six types of animals (three for each animal)

CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

served as targets (S2s). The list of the stimuli is shown in Table 1. The selection method of stimuli is documented in Appendix.

Table 4.1 A list of auditory and visual stimuli

Object	Auditory pre cue	Visual target (picture)
Dog	“woof”	
Chick	“Cheep”	
Cat	“meow”	
Crow	“caw”	
Sheep	“baa”	
Cow	“moo”	

Schematic Diagram of Visual Stimuli (S2)

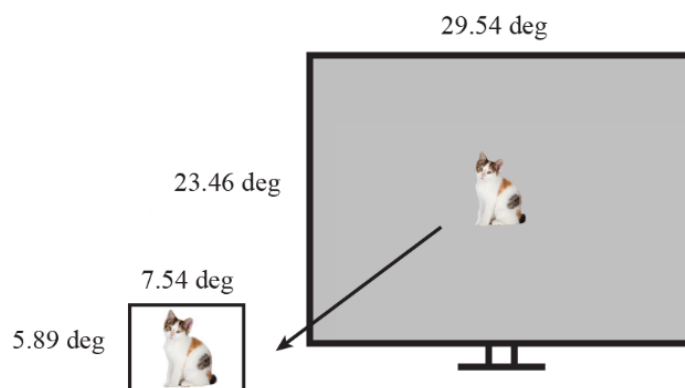


Fig. 4.1. shows the stimulus parameters of the present experimental stimuli. The visual target pictures that were embedded within the frame with 5.89° (visual angle) vertical and 7.54° horizontal were presented in the gray background with 23.46° vertical and 29.54° horizontal.

4.2.3 Procedure

As illustrated in Fig. 4.2, after the presentation of a fixation icon (300 ms), the prime was presented for 800 ms, and the target was presented for 400 ms successively with an interstimulus interval (ISI) of 1000 ms. Participants were asked to determine whether the target was semantically matched (congruent) as quickly as possible within 1300 ms after the target stimulus onset. The auditory-visual stimulus pairs were either semantically congruent or semantically incongruent, representing the same or different object categories, respectively. If any response was not given within the time window of 1300 ms, the next trial automatically began.

Illustration of the Experimental Design

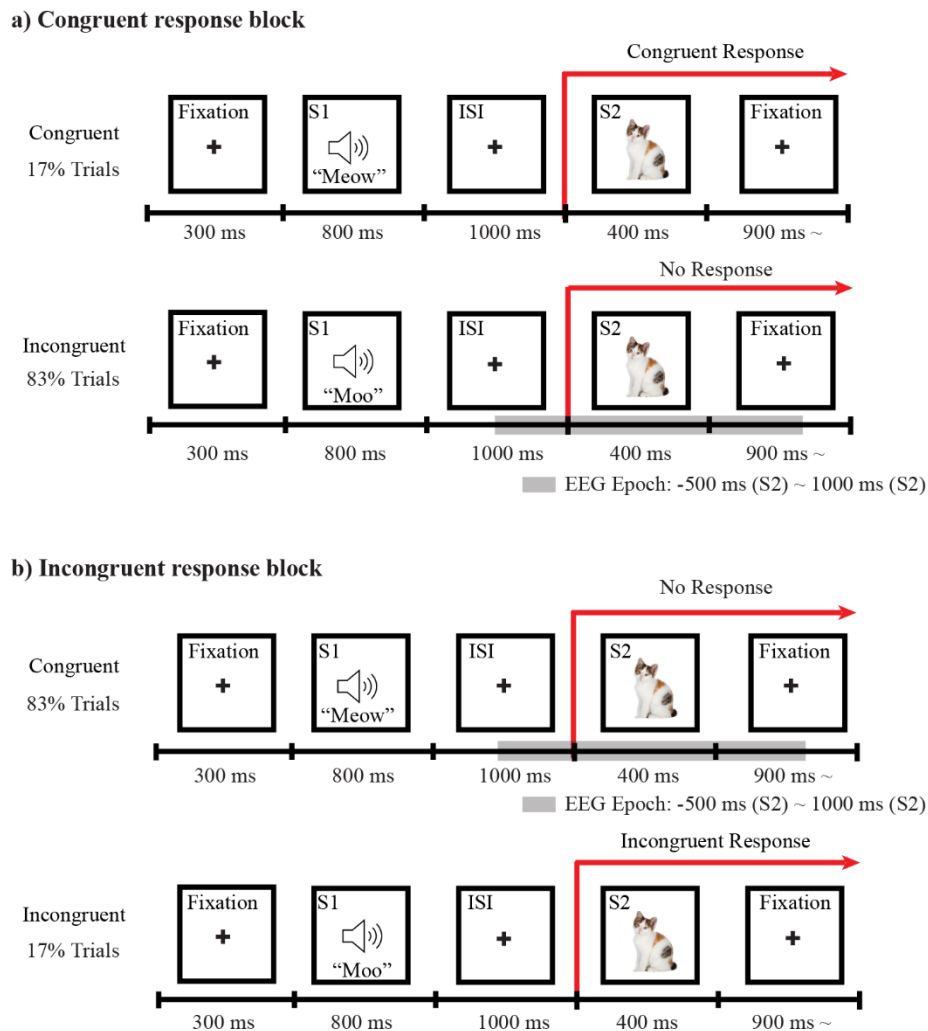


Fig. 4.2. An auditory prime (S1) and visual target (S2) were presented for 800 ms and 400 ms, respectively, following the presentation of a fixation icon for 300 ms. The interstimulus interval (ISI) was 1000 ms. The upper images represent a semantically congruent trial, and the lower images represent a semantically incongruent trial. Participants were asked to respond as fast and accurately as possible only after the second stimulus. The experimental session was divided into two response conditions: congruent and incongruent responses. Participants were required to perform three tasks, including an implicit response task. Under the congruent condition, participants were required to respond to the congruent pairs (same animal pairs), as

CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

a ‘congruent response’, and to the incongruent pairs (different animal pairs), with ‘no response’ (an implicit response). Under the incongruent condition, participants were required to respond to the incongruent pairs as an ‘incongruent response’ and to the congruent pairs with ‘no response’ (an implicit response). Participants’ ERPs were recorded during the no-response task to prevent possible effects of their hand movements.

The experimental session was divided into two kinds of response blocks: congruent and incongruent response blocks. Participants were required to perform three tasks, including an implicit response task. In the congruent response block, participants were required to respond to the congruent pairs (same animal pairs), as ‘congruent response’, and to the incongruent pairs (different animal pairs) with ‘no response’ (an implicit response). In the incongruent response block, participants were required to respond to the incongruent pairs as an ‘incongruent response’ and to the congruent pairs with ‘no response’ (an implicit response). EEG signals were recorded during the no-response outcome to prevent possible effects of participants’ hand movements.

In the congruent response block, the congruent and incongruent pairs were 17% and 83% of all pairs, respectively. In contrast, in the incongruent response block, the congruent and incongruent pairs accounted for 83% and 17% of all pairs, respectively. The experiment consisted of 20 sessions. Each session consisted of approximately 50 trials (46-54 trials). Behavioral data were collected from the participants’ congruent responses and incongruent responses. Participants’ ERP data were collected from the stimulus pairs during ‘no response’ tasks.

4.2.4 Data recordings

Continuous EEG data were collected from 29 scalp sites using sintered Ag/AgCl electrodes mounted on an elastic cap. Electrodes were positioned according to the international 10-20 standard (Jasper, 1958). All electrodes were referenced to the left earlobe, and the ground electrode was positioned on the forehead. Two electrodes were positioned below and on the sides of the eyes to record the EEG data. Skin-electrode contact impedance levels were maintained below 5 k Ω . EEG signals were recorded continuously at a sampling rate of 500 Hz.

4.2.5 Data analysis

Analysis of ERPs

ERPs were analyzed using Letswave and MATLAB 2014 (The MathWorks, Inc.) (Mouraux & Iannetti, 2008). (1) The obtained data were filtered in this step. The frequency of the low-pass filter was 1 Hz, and the frequency of the high-pass filter was 100 Hz. The notch filter was 60 Hz. (2) Ocular correction was conducted via a semiautomatic independent component analysis (ICA)-based correction process. (3) A new reference was used to select the electrode to be used for analysis (the online reference was the left earlobe, and the linked earlobes were exclusively calculated offline). (4) The obtained brain wave data were divided for the trigger point in each condition. Additionally, we set the start and end points of division. In this analysis, the point of 500 ms before the target stimulus was presented was defined as the starting

CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

point, and the point of 1000 ms after presentation was regarded as the end point. (5) We removed amplitudes with a minimum of $-80 \mu\text{V}$ and a maximum of $80 \mu\text{V}$ as artifacts. (6) The brain wave data obtained for each condition were averaged in this step. The potential reference value of the EEGs was calibrated and averaged over each condition. In our experiment, the average value of the potential from -200 ms to 0 ms at the moment of target stimulus presentation was calibrated to the reference value. Statistical analysis of the ERPs in response to visual stimuli was carried out for nine regions of interest (ROIs). Regions were defined as left frontal (LF: Fp1 and F7), middle frontal (MF: F3, Fz, and F4), right frontal (RF: Fp2 and F8), left central (LC: FC5, T7, C3, and CP5), middle central (MC: FC1, FC2, Cz, CP1, and CP2), right central (RC: FC6, T8, C4, and CP6), left posterior (LP: P3, P7, and O1), MP: (Pz, POz, and Oz), and right posterior (RP: P4, P8, and O2). Statistical analysis was performed using average data in each region that included two to five electrodes. To obtain information about the temporal evolution of significant differences, a cluster-based permutation test was performed to test differences between responses in the congruent and incongruent conditions at each ERP sampling point. We did not further analyze ERPs in response to auditory stimuli.

Analysis of source estimations

Analysis of source estimations used to find the difference brain activation area between congruent and incongruent conditions. We use an algorithm named

‘standardized low resolution electromagnetic tomography’ [35]. This is a method for estimating cortical generator localization using nonparametric statistical analysis (the method corrects for multiple comparisons and does not require the Gaussianity assumption). Localization was performed in 6239 cortical gray matter voxels of size 5 mm³ [88].

Analysis of gamma-band activity

Our analysis was carried out by a similar method as that used in previous studies [134, 135]. The method of time-frequency analysis was based on continuous wavelet transforms (CWTs), in which EEG signals were projected from the time domain to the time-frequency domain. Time–frequency analysis was performed for each channel by convolving the data with a complex Morlet wavelet. We used the mother wave as ‘Cwrt 1-3’ (Q increasing=3) for frequencies from 30 to 80 Hz (step size 2 Hz) in gamma-band activity analysis. The time-frequency analysis is based on a single-trial level and the analysis results in total power (with phase-locked and non-phase-locked signals). The baseline correction is based on the following equation: $P(t, f)_{\text{corrected}} = 100 \times \frac{(P(t, f) - P_{\text{baseline}}(f))}{P_{\text{baseline}}(f)}$. The period (-300 to -100 ms) before S2 served as baseline for all time-frequency analyses. Grand mean time–frequency results were computed over all subjects.

4.3 Results

4.3.1 Behavioral data

The behavioral results obtained from the data of both the congruent and incongruent responses for the 160 total trials are shown as the mean and standard deviation (SD) in Fig. 4.3. The accuracies for the congruent conditions (mean=94.07%, SD=4.28%) were almost the same ($p=0.087$) as those for the incongruent condition (mean=89.26%, SD=11.18%). The reaction times for the congruent condition (mean=516.8 ms, SD=68.87 ms) were faster ($p<0.0001$) than those for the incongruent condition (mean=606.3 ms, SD=73.1 ms).

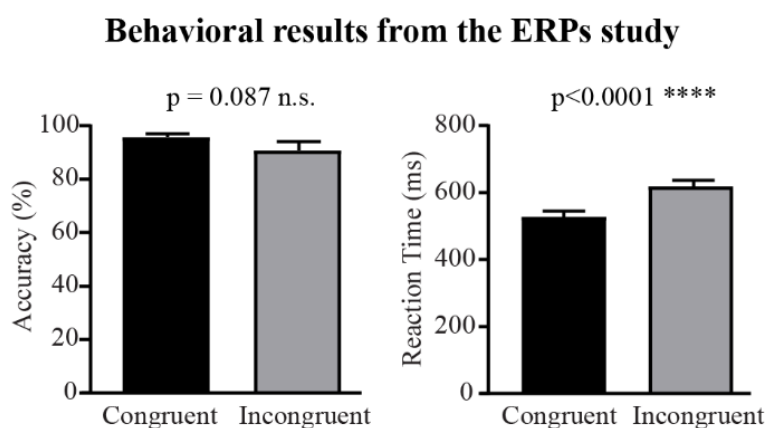


Fig. 4.3. Congruent responses had greater accuracy ($p=0.087$) and significantly shorter reaction times ($p<0.0001$) than incongruent responses.

4.3.2 Effects of semantic congruency on ERPs

The ERPs in response to visual targets following semantically congruent and

CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

semantically incongruent auditory primes and the difference in ERPs between the congruent and incongruent conditions are shown for the nine ROIs in Fig. 4.4. The ERPs for both congruent and incongruent conditions reveal an N1 component at 100 ms, a P1 component at 150 ms and an N2 component at 200 ms after visual stimulus onset, having similar characteristics in the time window of 0~200 ms. After 200 ms, the ERPs for the incongruent condition (blue trace) show a negative deflection compared to those for the congruent condition (red trace) between 250 ms and 350 ms. The difference in ERPs (black trace) show a negative peak at approximately 300 ms. A cluster-based permutation test (number of permutations = 2000) was performed using the Letswave toolbox with $\alpha = 0.05$ for cluster thresholding. The running cluster-based permutation test indicates significant differences between the conditions beginning at approximately 250 ms. The gray bar illustrates the time range with a significant difference.

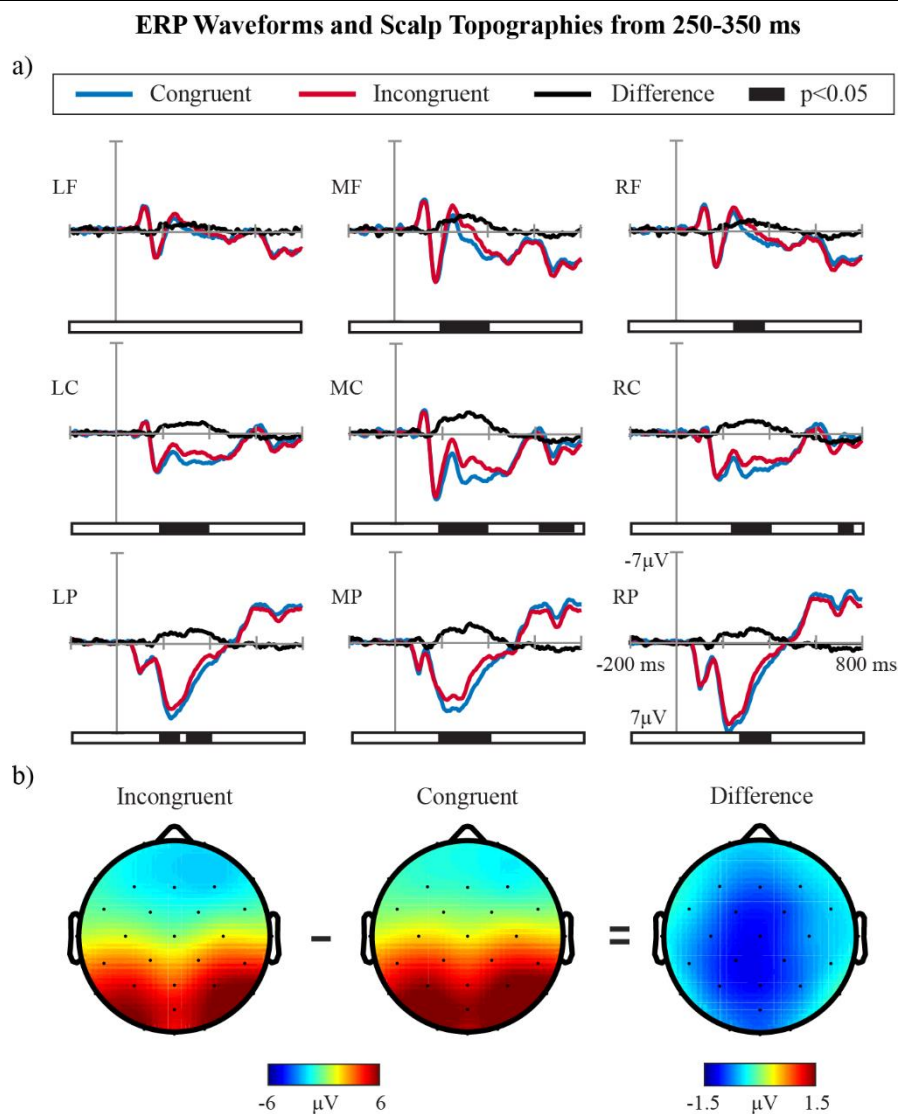


Fig. 4.4. ERPs and difference in ERPs with congruent pairs and incongruent pairs. ERPs in response to visual targets following semantically congruent (blue trace) and semantically incongruent auditory primes (red trace) are shown for the nine regions of interest. The difference in ERPs (black trace) between the congruent and incongruent conditions are also shown. A cluster-based permutation test (number of permutations = 2000) was performed using the Letswave toolbox with $\alpha = 0.05$ for cluster thresholding. The running cluster-based permutation test indicates significant differences between the conditions beginning at approximately 250 ms. The gray bar illustrates the time range of significant differences.

4.3.3 Source estimations of ERPs

The findings of the sLORETA analysis indicated that, the difference was statistically significant ($p < 0.05$) using a one-tailed t-test: Congruent condition $>$ Incongruent condition. A cluster-based permutation test (number of permutations = 5000) was performed using the sLORETA analysis toolbox with $\alpha = 0.05$ for cluster thresholding. Fig. 4.5 shows the differences in source localization between the congruent and incongruent conditions. Significant differences are marked in red. The results from the sLORETA imaging indicate decreased neuronal activation within the frontal lobes and the occipital lobes. In addition, we found significant differences in the lMFG and rSFG.

Multisensory Congruency Effects at 250-350 ms after Visual Stimulus Onset

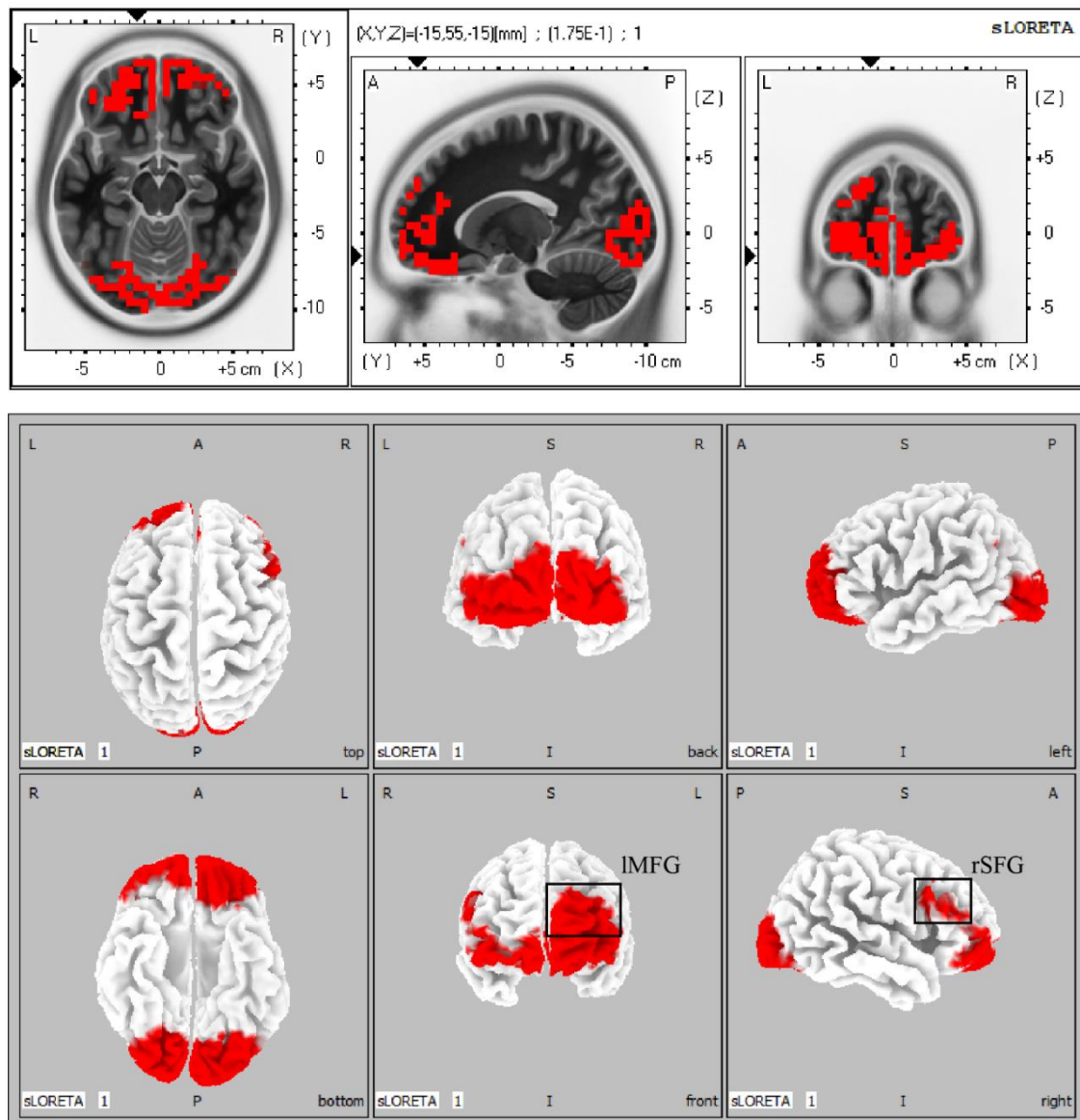
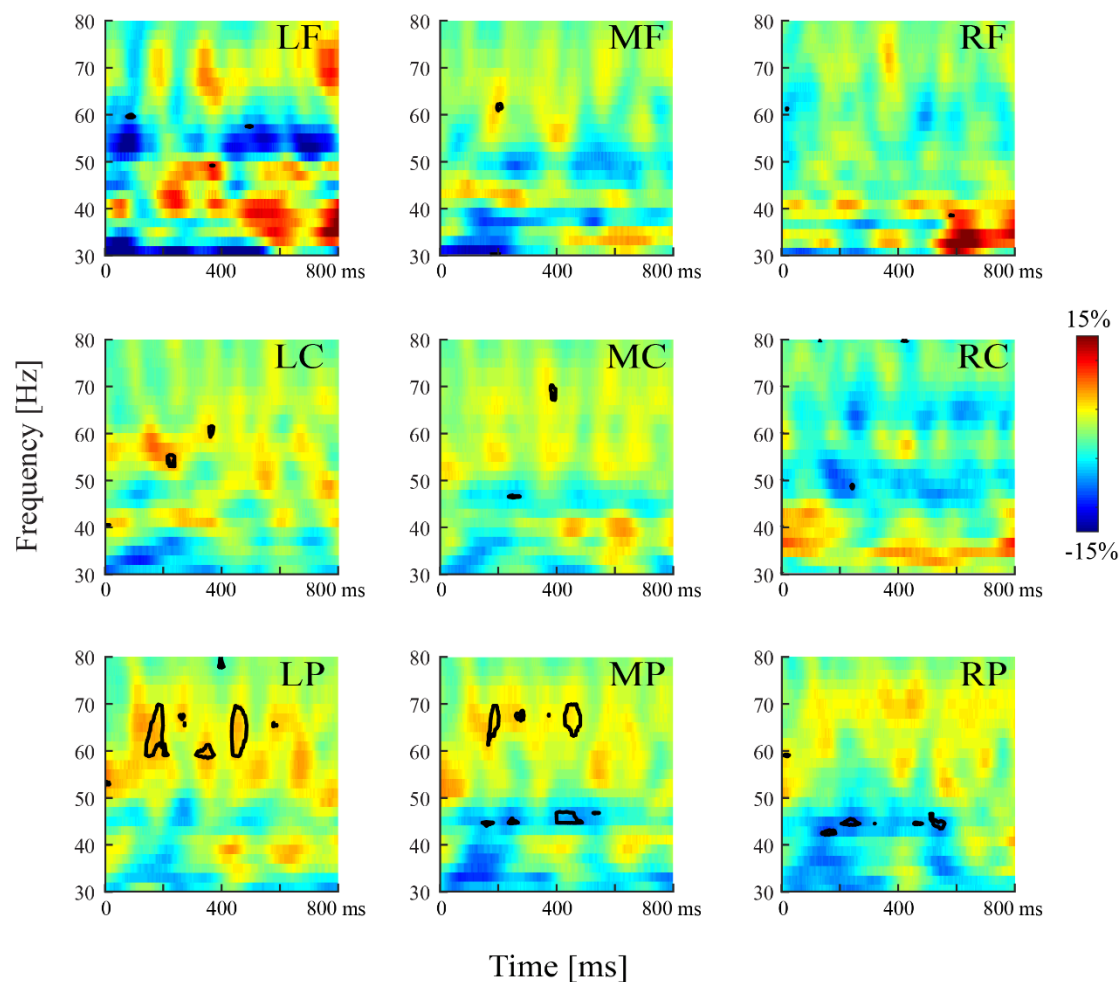


Fig. 4.5. Multisensory congruency effect at 250–350 ms. We compared the differences in source localization between the congruent and incongruent conditions. Significant differences are marked in red. Significant level is $p < 0.05$.

4.3.4 Effects of semantic congruency on gamma-band activity

Fig. 4.6 shows the time-frequency representation of the visual response for congruent, incongruent and difference at MC, LP and MP ROIs. A peak in gamma-band activity is present in the congruent and incongruent condition, reflecting an oscillatory neural response to visual stimulation with early and late latency. The time-frequency representations display total power. The contour line shows significant differences between the conditions ($p < 0.01$). As shown in Fig. 4.6, a significant difference of gamma oscillation between the congruent and incongruent conditions was observed in the frequency-time windows of approximately 60-70 Hz at 200 ms and 400 ms in the MC, LP and MP region.



CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

Fig. 4.6a. Time-frequency plots of total gamma-band activity in response to visual targets in 9 ROIs. The analysis was based on the difference between congruent and incongruent conditions. The plots show total oscillatory activity expressed as percent change relative to baseline following semantically congruent or incongruent primes. The black contour line in the plot to the right indicates significant differences ($p < 0.01$).

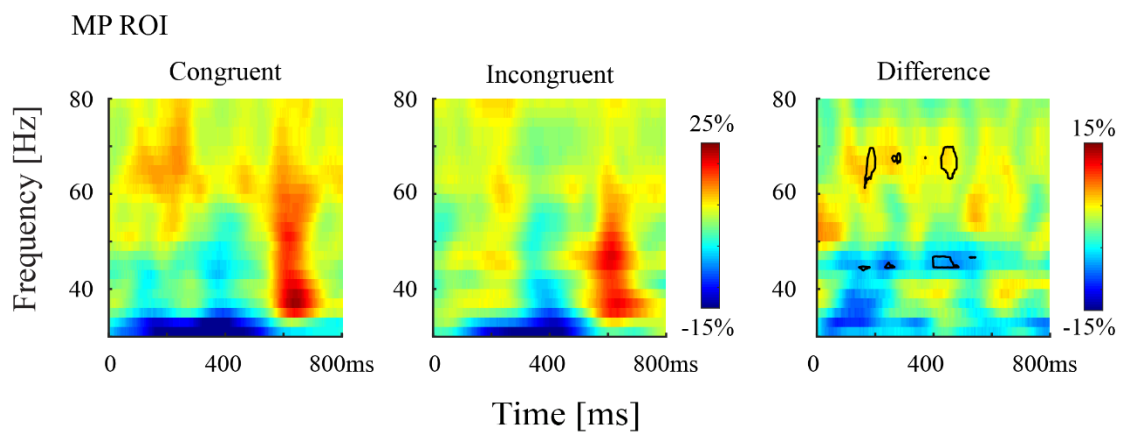


Fig. 4.6b. Time-frequency plots of total gamma-band activity in response to visual targets in MP ROI. The black contour line in the plot to the right indicates significant differences ($p < 0.01$).

CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

4.3.5 Results of ERPs classified by prime-target pairs

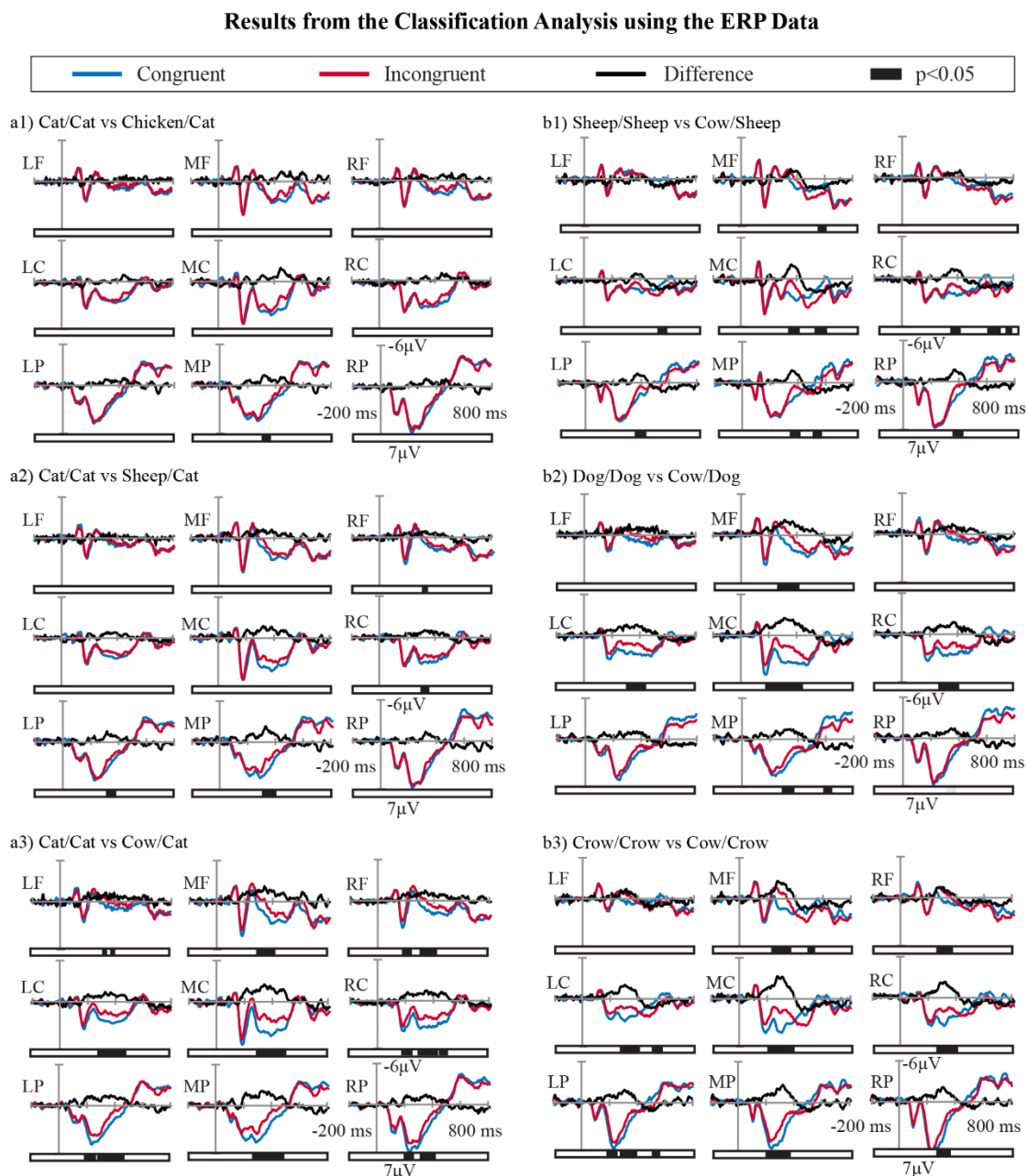


Fig. 4.7. ERPs and difference in ERPs with cat-target pairs and cow-prime pairs. In Fig. 4.7a and 4.7b ERPs in response to visual targets (cat) following semantically congruent (blue trace) and semantically incongruent auditory primes (red trace) are shown for the nine regions of interest. The difference in ERPs (black trace) between the congruent and incongruent conditions is also shown. A cluster-based permutation test (number of permutations = 2000)

CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

was performed using the Letswave toolbox with $\alpha = 0.05$ for cluster thresholding. The running cluster-based permutation test indicates significant differences between the conditions beginning at approximately 250 ms. The gray bar illustrates the time range of significant differences.

Fig. 4.7a shows the data for cat target based pairs in congruent and incongruent conditions. a1) cat/cat vs chicken/cat; a2) cat/cat vs sheep/cat; a3) cat/cat vs cow/cat. Fig. 4.7b shows the data for cow prime based pairs in congruent and incongruent conditions, b1) sheep/sheep vs cow/sheep; b2) crow/crow vs cow/crow; b3) dog/dog vs cow/dog. It is clearly shown that the difference ERPs depend on the stimulus type of prime-target pairs. For example, the differences in ERPs for cow/cat vs cat/cat are larger than those for chicken/cat vs cat/cat. The difference in ERPs for the cow/crow vs the crow/crow pair is larger than those for the cow/sheep vs the sheep/sheep pairs.

4.4.6 Correlation between ERP amplitude difference and semantic size difference

The purpose of this part was to verify the relationship between ERP amplitude differences and semantic size differences. We selected the average value of the different waveforms in the range of 200-400 ms as the measure of ERP amplitude difference in MC ROI, and the semantic size difference was obtained in evaluation tests (the size difference between prime animal and target animal). Then the data under each condition are sorted and analyzed by each pairing condition. Next, the calculated inverses of similarity (differences) led us to express the relationship between the differences in

CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

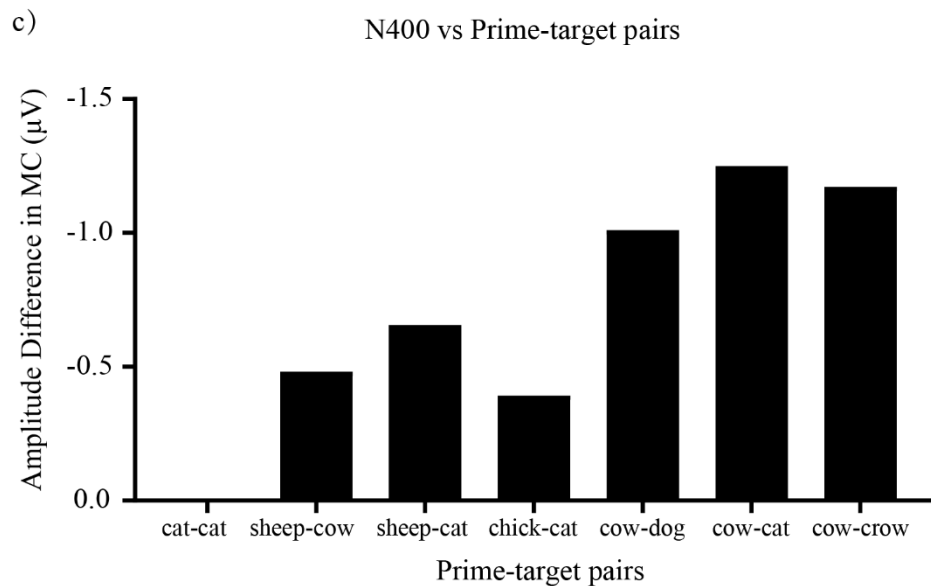
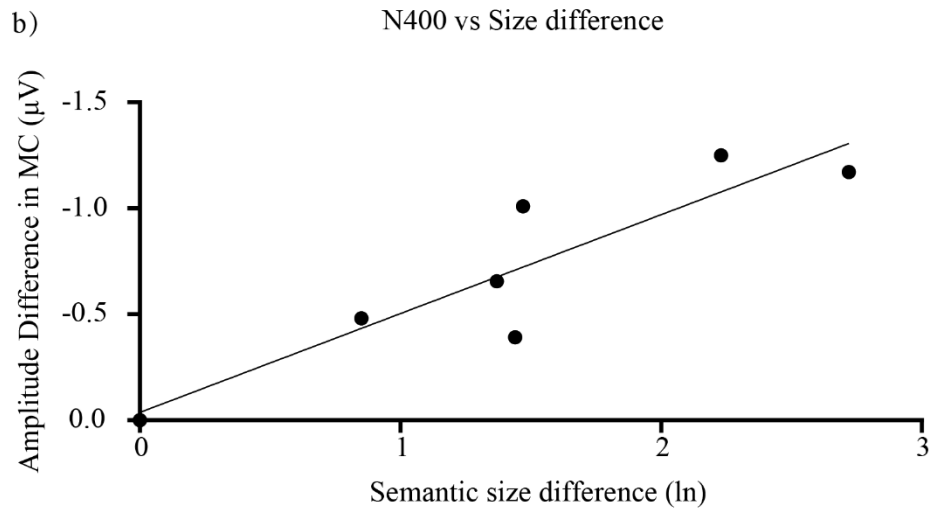
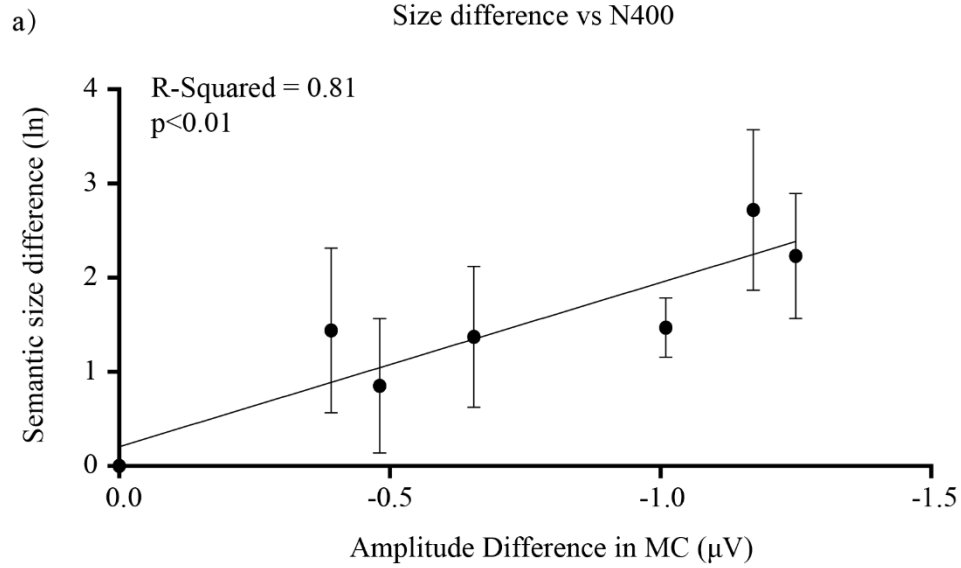
ERP amplitude and the differences in the semantic size difference of the individual animal pairs, as shown in Fig. 4.8b. It is clear that there is a simple linear functional relationship between differences in ERPs and the differences in the semantic size of the individual animal pairs. Further analysis of the data reveals a significant correlation using GraphPad Prism 8. The correlations for ERP amplitude differences and semantic size difference at the individual level are shown in Fig. 4.8, R-Squared = 0.81, $F(1,5)=21.73$, $p<0.01$ $y = -1.74 * x - 0.20$. This suggests that the differences in semantic size of individual animal pairs may cause the observed negative-going deflection effects in the ERPs; a larger difference in semantic size may yield larger negative-going deflections in ERPs.

Table 4.2 A list of evaluation semantic size

Name of animal	Semantic size (Ln)	Pair of animal	Difference of semantic size (Ln)
Cow	5.3840	Cow/Cat	2.2276
Sheep	4.5297	Cow/Crow	2.7203
Cat	3.1564	Cow/Dog	1.4720
Crow	2.6638	Sheep/Cat	1.3733
Chicken	1.7151	Chicken/Cat	1.4413
Dog	3.9120	Cow/Sheep	0.8543

We used a magnitude estimation method to examine visual similarity among the six types of animals (Table 4.2). The evaluation method is documented in Appendix.

CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION



CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

Fig. 4.8. Correlation between the average of ERP amplitude difference and semantic size.

Fig. 4.8 shows the ERP amplitude difference (MC) vs the semantic size difference. There is a significant relationship between the two differences. R-Squared = 0.81, $F(1,4)=21.73$, $p<0.01$

$$y = -1.74 * x - 0.20.$$

4.4 Discussion

4.4.1 Multisensory effects on event-related potentials

The crossmodal priming paradigm has been used to investigate object recognition across various modalities. In the present study, using an auditory-to-visual priming paradigm and the EEG method, we examined the effect of naturalistic sound on visual object categorization in semantically auditory-visual congruent and incongruent conditions. The ERPs for the incongruent condition showed a negative deflection compared to those for the congruent condition between 250 ms and 450 ms. The difference in ERPs showed a negative peak at approximately 300 ms.

These features are very similar to the N400, a component that was first found by Kutas and Hillyard and was thought to reflect contextual semantic integration processes [96, 136]. The N400 component was distinguished by the fact that it is a negative-going potential with its peak at approximately 400 ms poststimulus onset—between approximately 250 and 550 ms. Other researchers confirmed the semantic priming effect using a variety of tasks in unimodal and crossmodal conditions. Many previous studies are expected to converge towards the understanding of the semantic priming effect.

However, only a few researchers have paid attention to the priming effects for auditory prime and visual target stimuli, similar to the present study [84, 85, 130, 132]. Holcomb and Anderson found that the ERP priming effect of animal sounds to visual words was significant between 300 and 550 ms in stimulus-onset asynchrony (SOA)

CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

conditions of 200 and 800 ms [130]. Orgs et al. showed that environmental sounds (e.g., dog barking), instrument sounds (e.g., violin), everyday sounds (e.g., bell ringing) and visually displayed words produced an N400 effect between 200 and 500 ms in unrelated trials compared to related trials [132].

Brandman and Peelen showed that both natural sounds (e.g., animal sounds and spoken words, relative to uninformative noise) significantly facilitated visual object category MEG decoding between 300~500 ms after target (visual) onset. The time windows of 300~550 ms, 200~500 ms and 300~500 ms in the above studies are similar to those in the present study [85].

The relevant aspect of our study to the abovementioned studies is the high probability that labeling an auditory prime takes place because it is easy to identify the sources that produced them. If a prime stimulus is labeled, differences between related and unrelated items may reflect a linguistic, conceptual or crossmodal semantic relatedness effect associated with memory.

In this respect, not surprisingly, the time window of 250~450 ms observed in the present study is similar to those in 300~550 ms, 200~500 ms and 300~500 ms time windows in the above studies.

Regarding the topographic maps, the present difference in ERPs for the sheep/cat vs cat/cat, cow/cat vs cat/cat, cow/dog vs cow/cow and cow/crow vs crow/crow pairs showed a frontal-central distribution, as shown in Fig. 4.7. This mapping is similar to the results for word targets by Orgs et al [132]. but differs from the results for sound targets by Orgs et al. and Schneider, Debener, et al., which showed a central-posterior

CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

distribution [132, 134]. This suggests that crossmodal processing differs for visual to auditory tasks (auditory categorization) and auditory to visual tasks (visual categorization).

4.4.2 Source localization of multisensory effects

Prior studies have noted the importance of brain activation of crossmodal semantic matching. With respect to the first research question, a significant congruency effect was found in sLORETA. The most obvious finding to emerge from the analysis is the different neuronal activation within the frontal lobes and the occipital lobe. This finding of neuronal activation within the frontal lobes is consistent with that of Schneider, Debener, et al [134]. What is curious about this result is that different neuronal activation occurs within the occipital lobe. This result may be explained by the fact that the congruent effect enhances the visual response by promoting memory. Contrary to expectations, this study did not find a significant right lateralization. These findings may be somewhat limited by the methodological and the lower spatial resolution of sLORETA. However, we found significant differences in the IMFG and rSFG. The IMFG plays an important role in visual imagery and the rSFG plays an important role in working memory. The present study raises the possibility that auditory to visual congruency can enhance the representation of visual imagery and enhance working memory processing. These findings might help others to better understand the role of multisensory congruency in the delay matching paradigm.

4.4.3 Multisensory effects on gamma-band activity

A strong relationship between gamma-band activity and cross-modal binding has been reported in the literature. One of the aims of this study was to determine the gamma-band activity under auditory to visual priming. The results of this study show/indicate that the gamma band response caused by audio-visual priming is similar to audio-visual priming. These results further support the hypothesis of early stage binding in cross-modal priming. One unanticipated result was that gamma oscillated in the higher band (60-70 Hz). This band is higher than the result of Schneider, Debener, et al [134]. However, 30-70 Hz is the usual range of observations in visual research using gamma band analysis [137]. This difference may stem from differences in the sensory input channel. Another unanticipated result was that the negative gamma oscillatory in MP at 40 Hz. This result has not previously been described. Gamma-band activity and ERPs reflect, from different perspectives, differences between semantically congruent and incongruent conditions. ERPs differences may be due to different processes in the slow wave generator, reflecting phase-locking activity within low frequencies. The difference in gamma-band responses may respond to differences in early high-frequency oscillatory activity, reflecting phase-locked and non-phase-locked activity within high frequencies. This (rather) unexpected finding might be a result of the difference brain activation patterns between visual stimulation and auditory stimulation. Another source of uncertainty is the fewer electrodes and the choice of mother wavelet in offline analysis.

4.4.4 Multisensory incongruency is affected by the expected semantic size

Fig. 4.7 and Fig. 4.8 show that the difference in the ERP waveforms between congruent and incongruent conditions greatly depends on the priming-target pairs. Indeed, the difference decreases in the order of magnitude for the ‘cow/dog vs dog/dog’, ‘cow/cat vs the cat/cat’, ‘cow/crow vs crow/crow’, ‘cow/sheep vs sheep/sheep’, ‘sheep/cat vs cat/cat’ and ‘chicken/cat vs the cat/cat’ conditions.

Here, we examined the cause of the decreasing order of magnitude in the ERP difference. There are two possible routes by which auditory information could engage in the facilitation of visual objects [85]. One is the semantic route (indirect route), in which sound is first transformed into a semantic representation and then indirectly influences visual processing. The other is the direct route, in which sound directly activates the relevant object representations in the visual cortex. In this model, direct auditory-visual mappings could arise from the natural cooccurrence of visual objects with their corresponding visual object. On the other hand, semantic auditory-visual mappings could arise from a semantic representation with their relevant visual object.

If the semantic route was used in the present experiment, the semantic representation driven by a prime stimulus must have visual features of size, texture, shape and color similar to those of a target stimulus. On the other hand, the direct route could be used in the present experiment; in this case, the prime stimulus must have auditory features similar to those of a target sound.

To test the two model routes, by using a magnitude estimation method, we measured the semantic information driven by the auditory prime stimulus that was compatible

CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

with the visual target stimulus. The measurement was based on the hypothesis that the most important metric to judge congruency between an auditory prime and a visual target may be stimulus size because other object visual features such as texture, shape, and color are considered not to be compatible with one another.

Based on the hypothesis that dissimilarity (incongruency) might be determined predominantly by differences in stimuli, we estimated congruency responses for cat/cat, sheep/cat and cow/cat pairs. When the prime and target were both cats, the response for the same prime and target of the cow was congruent, as shown in fig. 4.9. The response for sheep/cat would be slightly incongruent because of the small difference in size between sheep and cats, while the response for cow/cat prime of cows would be largely incongruent because of the large difference in size between cows and cats.

The dependency could be accounted for by a hypothetical model in which the semantic information driven by the auditory prime and the information of the visual target may be integrated during visual object processing.

CHAPTER 4 THE INTERACTION PROCESSING BETWEEN AUDITION AND VISION

Model to Explain the Dependency of Different ERPs in Prime-Target Types

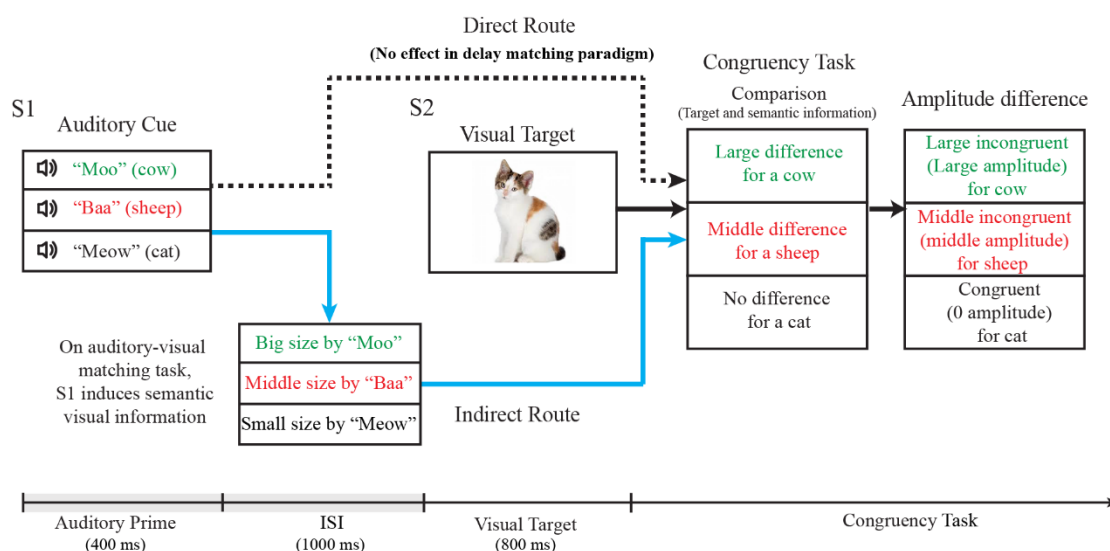


Fig. 4.9. Model to explain the dependency of different ERPs in prime-target types. Our model is based on the two-route model proposed by Brandman and Peelen [85]. Visual object processing receives visual information of the target in the direct route, creates a visual representation in the indirect route by an auditory prime, integrates both pieces of information, and outputs congruency depending on the similarity between the two pieces of information. Note here that auditory priming information in the direct route is not efficient for visual object processing. If the dissimilarity between the two pieces of information is larger, the incongruency is larger. We hypothesized that dissimilarity might be determined predominantly by differences in stimulus size. When the prime and target are cats, the response is congruent. Responses for the sheep/cat were incongruent with the expected size image, while responses for the cow/cat were incongruent with the unexpected size image.

4.5 Conclusion

The purpose of the current study was to determine the dependency on prime-target semantic size differences in an auditory-to-visual priming paradigm. This study has shown that auditory to visual priming is similar to visual to auditory priming in terms of the N400 effect, frontal lobe activation and early stage gamma-band activity. However, auditory to visual priming is not similar to visual to auditory priming regarding occipital lobe activation, as there are significant differences in lateralization in the left middle frontal gyrus (IMFG) and right superior frontal gyrus (rSFG) and a higher frequency-time window of gamma-band activity. For the first time, we present the difference of in peak amplitudes between congruent and incongruent conditions strongly dependent on the semantic size of prime-target pairs. The dependency suggests the congruency effects determined by the semantic size difference in auditory to visual priming paradigm. These findings could be accounted for by a hypothetical model in which the semantic information driven by the auditory prime and the information of the visual target may be integrated during visual object processing.

Chapter 5. Conclusion

In this study, we investigated the visual object recognition by manipulating object relations in multiple levels, such as flanking category, memory color, 3D depth scene and audio-visual crossmodality.

In chapter 2, we addressed the violation effect of symbol type under crowding in which we examined the spatial context effect at the high level of the visual processing. Our result of critical spacing showed that as the category's effect became stronger, the intensity of crowding was reduced in longer exposure time. Using the visual masking paradigm, we evaluated the category's effect in experiment 2. We proposed that the crowding at the high-level processing first increased until the peak value at about 145 ms and then decreased with SOA, suggesting that the crowding at high-level is similar to middle level during a specific time window.

In chapter 3, we addressed the violation effects of memory color and 3D depth scene in which we examined the temporal context effect at the high level of the visual processing. The significant difference between valid and invalid stimuli was observed at 450ms in color condition and was observed at 200ms and 400ms in depth condition. The significant difference between valid and invalid stimuli were found in color (425-480ms, Sub-Gyral) and depth condition (180-250ms, IPL; 425-480ms, MFG) by using sLORETA analysis. The significant difference between valid and invalid stimuli was mediated at theta and beta band in color condition, alpha and beta band in depth condition by using time frequency analysis. The significant difference was mediated at

inferior frontal gyrus (IFG) in theta band (4–7 Hz) and superior parietal lobule (SPL) in beta2 band (16.5–20 Hz) under color condition. Violation effect occurred in different time window and different brain areas among memory color and 3d depth scene, suggesting that the violation of color memory and 3d depth scene are mediated by different brain mechanism.

In chapter 4, we addressed auditory-visual violation effect in which we examined the temporal context effect at the high level of the auditory-visual crossmodal processing. Our result showed that the auditory to visual priming was similar to the visual to auditory priming about N400 effect, frontal lobe activation and the early stage gamma-band activity in the previous study. However, the auditory to visual priming was not similar to the visual to auditory priming about the occipital lobe activation, the significant differences of lateralization in left middle frontal gyrus (lMFG) and right superior frontal gyrus (rSFG), and the higher frequency-time window of gamma-band activity in the previous study. This finding is different from the finding for visual-to-auditory by Schneider et al. This implies that auditory-to-visual crossmodal processing is different from visual-to-auditory processing. In additional, the N400 effect was strongly dependent on the difference of semantic size between prime-target pairs. The dependency could be accounted for by a hypothetical model in which the semantic information driven by the auditory prime and the information of the visual target may be integrated during visual object processing, suggesting that the congruency or incongruency between an auditory prime and a visual target may be mediated by the interaction of bottom-up and top-down systems.

Acknowledgments

Firstly, I would like to express my sincerely gratitude to my supervisor Prof. Jinglong Wu for continuous supports in my Ph.D. studies and related researches. His diligence gives me a good example, which enables me to have a good understanding of my working attitude during my Ph.D.

Secondly, I would like to express my sincerely gratitude to Prof. Satoshi Takahashi. He gave me a lot of comments when I made my study plan, conducted my experiments, and wrote the paper.

Thirdly, I would like to express my sincerely thank to Assistant Professor Jiajia Yang, who provided me a lot of comments when I wrote my paper. He gave me precious supports to help me complete this thesis successfully.

At last, I wish to express my gratitude to Prof. Yoshimichi Ejima for his constant support and constant encouragement. Thank you for your guidance on my thesis and the knowledge you taught me. I cannot forget this life.

Appendix

Evaluation of experimental stimuli for Chapter 4

Twenty-eight healthy volunteers (5 female) participated in the first test. We selected six types of animals from a set of 18 types of animals (dog, horse, cat, lion, elephant, cow, sheep, pig, mouse, chicken, monkey, crow, chick, sparrow, duck, rabbit, panda, bear and frog) that were considered suitable for the categorization task. Animals were selected by a questionnaire test assessing the characteristics of animal bark sounds with familiarity, uniqueness and reliability for categorization. The selected animals were dogs, cats, cows, chickens, sheep and crows.

Twenty healthy volunteers (3 female) participated in the second test. We created an auditory stimulus for each animal that was to be the most relevant to each animal sound through a psychophysical experiment in which participants' accuracy and confidence were assessed for six sounds of each animal; the results for each animal were compared with each other. The auditory stimuli with the highest relevance were used.

Twenty healthy volunteers (5 female) participated in the third and fourth tests. We created three visual stimuli for each animal by selecting those with the best fit to the auditory stimulus of each animal and the well discriminating pictures from the auditory stimulus of the other animals. Selections were made through a psychophysical experiment in which auditory-visual congruency was ranked from levels 1 to 7 for various combinations of auditory and visual stimuli. The levels of '7' and '1' indicate the highest congruency and lowest congruency (highest incongruency), respectively. In

this way, we selected six kinds of auditory stimuli (dog × 1, chicken × 1, cat × 1, cow × 1, crow × 1, sheep × 1) and eighteen kinds of visual stimuli (dog × 3, chicken × 3, cat × 3, cow × 3, crow × 3, sheep × 3). The six auditory and eighteen visual stimuli (3 for each animal) selected yielded a high congruency (average of 6.18) between the same animal and a low congruency (average of 1.15) between different animals. These congruency values indicated that the selected stimuli were suitable for examining the crossmodal priming effects using auditory-visual events.

Ten healthy volunteers (5 female) participated in the fifth test. We used a magnitude estimation method to examine visual similarity among the six types of animals. Participants were required to estimate the visual size of the semantic stimulus driven by an animal sound by assigning numerical values proportional to the visual size of the standard animal (dog). The visual size of the standard animal (dog) was assigned a numerical value of 50. Using the data, we calculated the inverse of similarity between animal bark sounds, that is, the difference between them.

Reference

- [1] Schwartz O, Hsu A, Dayan P. Space and time in visual context. *Nature Reviews Neuroscience*. 2007;8(7):522-35.
- [2] Soranzo A. Simultaneous contrast, simultaneous brightness contrast, simultaneous color. 2016.
- [3] Herzog MH, Sayim B, Chicherov V, Manassi M. Crowding, grouping, and object recognition: A matter of appearance. *Journal of vision*. 2015;15(6):5-.
- [4] Polat U, Sagi D. Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision research*. 1993;33(7):993-9.
- [5] Gibson JJ. Adaptation, after-effect, and contrast in the perception of tilted lines. II. Simultaneous contrast and the areal restriction of the after-effect. *Journal of Experimental Psychology*. 1937;20(6):553.
- [6] Reuther J, Chakravarthi R. Categorical membership modulates crowding: Evidence from characters. *Journal of Vision*. 2014;14(6):5-.
- [7] Fischer J, Whitney D. Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology*. 2011;106(3):1389-98.
- [8] Levi DM. Crowding—An essential bottleneck for object recognition: A mini-review. *Vision research*. 2008;48(5):635-54.
- [9] Pelli DG, Tillman KA. The uncrowded window of object recognition. *Nature neuroscience*. 2008;11(10):1129-35.

REFERENCE

- [10] Manassi M, Whitney D. Multi-level crowding and the paradox of object recognition in clutter. *Current Biology*. 2018;28(3):R127-R33.
- [11] Laparra V, Malo J. Visual aftereffects and sensory nonlinearities from a single statistical framework. *Frontiers in human neuroscience*. 2015;9:557.
- [12] Gibson JJ, Radner M. Adaptation, after-effect and contrast in the perception of tilted lines. I. Quantitative studies. *Journal of experimental psychology*. 1937;20(5):453.
- [13] Hill WE. My wife and my mother-in-law. *Puck*. 1915;16:11.
- [14] Gregory RL. The Medawar lecture 2001 knowledge for vision: Vision for knowledge. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2005;360(1458):1231-51.
- [15] Blakemore C, Carpenter RH, Georgeson MA. Lateral inhibition between orientation detectors in the human visual system. *Nature*. 1970;228(5266):37-9.
- [16] Brainard DH, Wandell BA. Analysis of the retinex theory of color vision. *JOSA A*. 1986;3(10):1651-61.
- [17] Land E. Recent advances in retinex theory *Vision Research*. 1986.
- [18] Wandell BA, Dumoulin SO, Brewer AA. Visual field maps in human cortex. *Neuron*. 2007;56(2):366-83.
- [19] Ungerleider LG. Two cortical visual systems. *Analysis of visual behavior*. 1982:549-86.

REFERENCE

- [20] Perry CJ, Fallah M. Feature integration and object representations along the dorsal stream visual hierarchy. *Frontiers in Computational Neuroscience*. 2014;8(84).
- [21] Van der Helm PA. *Simplicity in vision: A multidisciplinary account of perceptual organization*: Cambridge University Press; 2014.
- [22] Groen II, Silson EH, Baker CI. Contributions of low-and high-level properties to neural processing of visual scenes in the human brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2017;372(1714):20160102.
- [23] Humphreys GW, Price CJ, Riddoch MJ. From objects to names: A cognitive neuroscience approach. *Psychological research*. 1999;62(2):118-30.
- [24] Ward J. *The student's guide to cognitive neuroscience*: Routledge; 2019.
- [25] Teitelbaum RC, Biederman I, editors. *Perceiving Real World Scenes: The Role of a Prior Glance*. *Proceedings of the Human Factors Society Annual Meeting*; 1979: SAGE Publications Sage CA: Los Angeles, CA.
- [26] Wenger MJ, Gibson BS. Using hazard functions to assess changes in processing capacity in an attentional cuing paradigm. *Journal of Experimental Psychology: Human Perception and Performance*. 2004;30(4):708.
- [27] Watson AB, Pelli DG. QUEST: A Bayesian adaptive psychometric method. *Perception & psychophysics*. 1983;33(2):113-20.
- [28] Kleiner M, Brainard D, Pelli D. What's new in Psychtoolbox-3? 2007.

REFERENCE

- [29] Lesmes L, Lu Z-L, Baek J, Tran N, Doshier B, Albright T. Developing Bayesian adaptive methods for estimating sensitivity thresholds (d') in Yes-No and forced-choice tasks. *Frontiers in Psychology*. 2015;6(1070).
- [30] Macknik SL. Visual masking approaches to visual awareness. *Progress in brain research*. 2006;155:177-215.
- [31] Breitmeyer B, Ogmen H, Ögmen H. *Visual masking: Time slices through conscious and unconscious vision*: Oxford University Press; 2006.
- [32] Lindsley DB. *Psychological phenomena and the electroencephalogram*. *Electroencephalography & Clinical Neurophysiology*. 1952.
- [33] Luck SJ. *An introduction to the event-related potential technique*: MIT press; 2014.
- [34] Torrence C, Compo GP. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*. 1998;79(1):61-78.
- [35] Pascual-Marqui RD. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find Exp Clin Pharmacol*. 2002;24(Suppl D):5-12.
- [36] Bouma H. Visual interference in the parafoveal recognition of initial and final letters of words. *Vision research*. 1973;13(4):767-82.
- [37] Huckauf A, Heller D. On the relations between crowding and visual masking. *Perception & Psychophysics*. 2004;66(4):584-95.

REFERENCE

- [38] Pelli DG, Palomares M, Majaj NJ. Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of vision*. 2004;4(12):12-.
- [39] Whitney D, Levi DM. Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in cognitive sciences*. 2011;15(4):160-8.
- [40] Doerig A, Bornet A, Rosenholtz R, Francis G, Clarke AM, Herzog MH. Beyond Bouma's window: How to explain global aspects of crowding? *PLoS computational biology*. 2019;15(5):e1006580.
- [41] Millin R, Arman AC, Chung ST, Tjan BS. Visual crowding in V1. *Cerebral Cortex*. 2014;24(12):3107-15.
- [42] Pöder E, Wagemans J. Crowding with conjunctions of simple features. *Journal of Vision*. 2007;7(2):23-.
- [43] Nandy AS, Tjan BS. The nature of letter crowding as revealed by first-and second-order classification images. *Journal of Vision*. 2007;7(2):5-.
- [44] Hanus D, Vul E. Quantifying error distributions in crowding. *Journal of Vision*. 2013;13(4):17-.
- [45] Chakravarthi R, Cavanagh P. Temporal properties of the polarity advantage effect in crowding. *Journal of Vision*. 2007;7(2):11-.
- [46] Lev M, Polat U. Space and time in masking and crowding. *Journal of vision*. 2015;15(13):10-.

REFERENCE

- [47] Yeshurun Y, Rashal E, Tkacz-Domb S. Temporal crowding and its interplay with spatial crowding. *Journal of vision*. 2015;15(3):11-.
- [48] Chung ST. Spatio-temporal properties of letter crowding. *Journal of Vision*. 2016;16(6):8-.
- [49] Brainard DH. The psychophysics toolbox. *Spatial vision*. 1997;10(4):433-6.
- [50] Riddoch MJ, Humphreys GW. Object recognition. *The handbook of cognitive neuropsychology*. 2001:45-74.
- [51] Kafaligonul H, Breitmeyer BG, Ögmen H. Feedforward and feedback processes in vision. *Frontiers in psychology*. 2015;6:279.
- [52] Wutz A, Melcher D. The temporal window of individuation limits visual capacity. *Frontiers in psychology*. 2014;5:952.
- [53] Biederman I. Perceiving real-world scenes. *Science*. 1972;177(4043):77-80.
- [54] Biederman I, Glass AL, Stacy EW. Searching for objects in real-world scenes. *Journal of experimental psychology*. 1973;97(1):22.
- [55] Biederman I, Rabinowitz JC, Glass AL, Stacy EW. On the information extracted from a glance at a scene. *Journal of experimental psychology*. 1974;103(3):597.
- [56] Bar M. Visual objects in context. *Nature Reviews Neuroscience*. 2004;5(8):617-29.
- [57] Biederman I, Mezzanotte RJ, Rabinowitz JC. Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive psychology*. 1982;14(2):143-77.

REFERENCE

- [58] Chun MM. Contextual cueing of visual attention. *Trends in cognitive sciences*. 2000;4(5):170-8.
- [59] Ganis G, Kutas M. An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*. 2003;16(2):123-44.
- [60] Oliva A, Torralba A. The role of context in object recognition. *Trends in cognitive sciences*. 2007;11(12):520-7.
- [61] Palmer tE. The effects of contextual scenes on the identification of objects. *Memory & cognition*. 1975;3:519-26.
- [62] Davenport JL. Consistency effects between objects in scenes. *Memory & Cognition*. 2007;35(3):393-401.
- [63] Davenport JL, Potter MC. Scene consistency in object and background perception. *Psychological science*. 2004;15(8):559-64.
- [64] Fabre-Thorpe M. The characteristics and limits of rapid visual categorization. *Frontiers in psychology*. 2011;2:243.
- [65] Fize D, Cauchoix M, Fabre-Thorpe M. Humans and monkeys share visual representations. *Proceedings of the National Academy of Sciences*. 2011;108(18):7635-40.
- [66] Green C, Hummel JE. Familiar interacting object pairs are perceptually grouped. *Journal of Experimental Psychology: Human Perception and Performance*. 2006;32(5):1107.

REFERENCE

- [67] Gronau N, Neta M, Bar M. Integrated contextual representation for objects' identities and their locations. *Journal of cognitive neuroscience*. 2008;20(3):371-88.
- [68] Joubert OR, Fize D, Rousselet GA, Fabre-Thorpe M. Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*. 2008;8(13):11-.
- [69] Joubert OR, Rousselet GA, Fize D, Fabre-Thorpe M. Processing scene context: Fast categorization and object interference. *Vision research*. 2007;47(26):3286-97.
- [70] Kret ME, de Gelder B. Social context influences recognition of bodily expressions. *Experimental Brain Research*. 2010;203(1):169-80.
- [71] Mudrik L, Lamy D, Deouell LY. ERP evidence for context congruity effects during simultaneous object–scene processing. *Neuropsychologia*. 2010;48(2):507-17.
- [72] Hindy NC, Ng FY, Turk-Browne NB. Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature neuroscience*. 2016;19(5):665-7.
- [73] Kok P, Failing MF, de Lange FP. Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of cognitive neuroscience*. 2014;26(7):1546-54.

REFERENCE

- [74] Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L. Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience*. 2010;30(8):2960-6.
- [75] den Ouden HE, Daunizeau J, Roiser J, Friston KJ, Stephan KE. Striatal prediction error modulates cortical coupling. *Journal of Neuroscience*. 2010;30(9):3210-9.
- [76] Kok P, Jehee JF, De Lange FP. Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*. 2012;75(2):265-70.
- [77] Summerfield C, Trittschuh EH, Monti JM, Mesulam M-M, Egnér T. Neural repetition suppression reflects fulfilled perceptual expectations. *Nature neuroscience*. 2008;11(9):1004-6.
- [78] Todorovic A, de Lange FP. Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *Journal of Neuroscience*. 2012;32(39):13389-95.
- [79] Kok P, Turk-Browne NB. Associative prediction of visual shape in the hippocampus. *Journal of Neuroscience*. 2018;38(31):6888-99.
- [80] Bar M, Aminoff E. Cortical analysis of visual context. *Neuron*. 2003;38(2):347-58.
- [81] Bar M. A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of cognitive neuroscience*. 2003;15(4):600-9.

REFERENCE

- [82] Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, et al. Top-down facilitation of visual recognition. *Proceedings of the national academy of sciences*. 2006;103(2):449-54.
- [83] Livne T, Bar M. Cortical integration of contextual information across objects. *Journal of cognitive neuroscience*. 2016;28(7):948-58.
- [84] Brandman T, Peelen MV. Interaction between scene and object processing revealed by human fMRI and MEG decoding. *Journal of Neuroscience*. 2017;37(32):7700-10.
- [85] Brandman T, Avancini C, Leticevscaia O, Peelen MV. Auditory and semantic cues facilitate decoding of visual object category in MEG. *Cerebral Cortex*. 2020;30(2):597-606.
- [86] Caplette L, Ince RA, Jerbi K, Gosselin F. Disentangling presentation and processing times in the brain. *NeuroImage*. 2020;218:116994.
- [87] Wu Q, Wu J, Takahashi S, Huang Q, Sun H, Guo Q, et al. Modes of effective connectivity within cortical pathways are distinguished for different categories of visual context: an fMRI study. *Frontiers in behavioral neuroscience*. 2017;11:64.
- [88] Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*. 2002;15(1):1-25.

REFERENCE

- [89] Love J, Selker R, Marsman M, Jamil T, Dropmann D, Verhagen J, et al. JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*. 2019;88(1):1-17.
- [90] Olkkonen M, Hansen T, Gegenfurtner KR. Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *Journal of vision*. 2008;8(5):13-.
- [91] Wong NH, Ban H, Chang DH. Human Depth Sensitivity Is Affected by Object Plausibility. *Journal of cognitive neuroscience*. 2020;32(2):338-52.
- [92] Sutton S, Braren M, Zubin J, John E. Evoked-potential correlates of stimulus uncertainty. *Science*. 1965;150(3700):1187-8.
- [93] Folstein JR, Van Petten C. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology*. 2008;45(1):152-70.
- [94] Patel SH, Azzam PN. Characterization of N200 and P300: selected studies of the event-related potential. *International journal of medical sciences*. 2005;2(4):147.
- [95] Minami T, Goto K, Kitazaki M, Nakauchi S. Asymmetry of P3 amplitude during oddball tasks reflects the unnaturalness of visual stimuli. *Neuroreport*. 2009;20(16):1471-6.
- [96] Kutas M, Hillyard SA. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*. 1980;207(4427):203-5.

REFERENCE

- [97] Levi-Aharoni H, Shriki O, Tishby N. Surprise response as a probe for compressed memory states. *PLoS computational biology*. 2020;16(2):e1007065.
- [98] Shin J. The interrelationship between movement and cognition: Theta rhythm and the P300 event-related potential. *Hippocampus*. 2011;21(7):744-52.
- [99] Wang X, Ding M. Relation between P300 and event-related theta-band synchronization: A single-trial analysis. *Clinical Neurophysiology*. 2011;122(5):916-24.
- [100] Siegel M, Engel AK, Donner TH. Cortical network dynamics of perceptual decision-making in the human brain. *Frontiers in human neuroscience*. 2011;5:21.
- [101] Klimesch W, Sauseng P, Hanslmayr S. EEG alpha oscillations: the inhibition-timing hypothesis. *Brain research reviews*. 2007;53(1):63-88.
- [102] Klimesch W. Alpha-band oscillations, attention, and controlled access to stored information. *Trends in cognitive sciences*. 2012;16(12):606-17.
- [103] Klimesch W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*. 1999;29(2-3):169-95.
- [104] Kiss I, Dashieff RM, Lordeon P. A parietooccipital generator for P300: Evidence from human intracranial recordings. *International Journal of Neuroscience*. 1989;49(1-2):133-9.

REFERENCE

- [105] Adelhöfer N, Beste C. Pre-trial theta band activity in the ventromedial prefrontal cortex correlates with inhibition-related theta band activity in the right inferior frontal cortex. *Neuroimage*. 2020;219:117052.
- [106] Hsieh L-T, Ranganath C. Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. *Neuroimage*. 2014;85:721-9.
- [107] Osipova D, Takashima A, Oostenveld R, Fernández G, Maris E, Jensen O. Theta and gamma oscillations predict encoding and retrieval of declarative memory. *Journal of neuroscience*. 2006;26(28):7523-31.
- [108] Chmielewski WX, Mückschel M, Stock A-K, Beste C. The impact of mental workload on inhibitory control subprocesses. *NeuroImage*. 2015;112:96-104.
- [109] Johns P. *Clinical Neuroscience E-Book*: Elsevier Health Sciences; 2014.
- [110] Li F, Tao Q, Peng W, Zhang T, Si Y, Zhang Y, et al. Inter-subject P300 variability relates to the efficiency of brain networks reconfigured from resting-to task-state: evidence from a simultaneous event-related EEG-fMRI study. *NeuroImage*. 2020;205:116285.
- [111] Sperber RD, McCauley C, Ragain RD, Weil CM. Semantic priming effects on picture and word processing. *Memory & Cognition*. 1979;7(5):339-45.
- [112] Schneider TR, Engel AK, Debener S. Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Experimental psychology*. 2008;55(2):121-32.

REFERENCE

- [113] Chen Y-C, Spence C. The time-course of the cross-modal semantic modulation of visual picture processing by naturalistic sounds and spoken words. *Multisensory Research*. 2013;26(4):371-86.
- [114] Chen Y-C, Spence C. Dissociating the time courses of the cross-modal semantic priming effects elicited by naturalistic sounds and spoken words. *Psychonomic bulletin & review*. 2018;25(3):1138-46.
- [115] Edmiston P, Lupyan G. What makes words special? Words as unmotivated cues. *Cognition*. 2015;143:93-100.
- [116] Iordanescu L, Grabowecky M, Franconeri S, Theeuwes J, Suzuki S. Characteristic sounds make you look at target objects more quickly. *Attention, Perception, & Psychophysics*. 2010;72(7):1736-41.
- [117] Kim Y, Porter AM, Goolkasian P. Conceptual priming with pictures and environmental sounds. *Acta psychologica*. 2014;146:73-83.
- [118] Vallet GT, Simard M, Versace R, Mazza S. The perceptual nature of audiovisual interactions for semantic knowledge in young and elderly adults. *Acta psychologica*. 2013;143(3):253-60.
- [119] Bizley JK, Cohen YE. The what, where and how of auditory-object perception. *Nature Reviews Neuroscience*. 2013;14(10):693-707.
- [120] Chu S, Narayanan S, Kuo C-CJ. Environmental sound recognition with time–frequency audio features. *IEEE Transactions on Audio, Speech, and Language Processing*. 2009;17(6):1142-58.

REFERENCE

- [121] Cummings A, Čeponienė R, Koyama A, Saygin AP, Townsend J, Dick F. Auditory semantic networks for words and natural sounds. *Brain research*. 2006;1115(1):92-107.
- [122] Dudschig C, Mackenzie IG, Leuthold H, Kaup B. Environmental sound priming: Does negation modify N400 cross-modal priming effects? *Psychonomic Bulletin & Review*. 2018;25(4):1441-8.
- [123] Ferris TK, Sarter NB. Cross-modal links among vision, audition, and touch in complex environments. *Human Factors*. 2008;50(1):17-26.
- [124] Frey A, Aramaki M, Besson M. Conceptual priming for realistic auditory scenes and for auditory words. *Brain and cognition*. 2014;84(1):141-52.
- [125] Frings C, Schneider KK, Fox E. The negative priming paradigm: An update and implications for selective attention. *Psychonomic bulletin & review*. 2015;22(6):1577-97.
- [126] Hendrickson K, Love T, Walenski M, Friend M. The organization of words and environmental sounds in the second year: Behavioral and electrophysiological evidence. *Developmental science*. 2019;22(1):e12746.
- [127] Holcomb PJ, Neville HJ. Natural speech processing: An analysis using event-related brain potentials. *Psychobiology*. 1991;19(4):286-300.
- [128] Noppeney U, Lee HL. Causal inference and temporal predictions in audiovisual perception of speech and music. *Ann NY Acad Sci*. 2018;1423:102-16.

REFERENCE

- [129] Brandman, Peelen MV. Auditory and semantic cues facilitate decoding of visual object category in MEG. *Cerebral Cortex*. 2020;30(2):597-606.
- [130] Holcomb PJ, Anderson JE. Cross-modal semantic priming: A time-course analysis using event-related brain potentials. *Language and cognitive processes*. 1993;8(4):379-411.
- [131] Mushtaq Z, Su S-F. Efficient Classification of Environmental Sounds through Multiple Features Aggregation and Data Enhancement Techniques for Spectrogram Images. *Symmetry*. 2020;12(11):1822.
- [132] Orgs G, Lange K, Dombrowski J-H, Heil M. Conceptual priming for environmental sounds and words: An ERP study. *Brain and cognition*. 2006;62(3):267-72.
- [133] Van Petten C, Rheinfelder H. Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*. 1995;33(4):485-508.
- [134] Schneider TR, Debener S, Oostenveld R, Engel AK. Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *Neuroimage*. 2008;42(3):1244-54.
- [135] Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J. Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *Journal of Neuroscience*. 1996;16(13):4240-9.
- [136] Kutas M, Federmeier KD. Electrophysiology reveals semantic memory use in language comprehension. *Trends in cognitive sciences*. 2000;4(12):463-70.

REFERENCE

- [137] Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J. Oscillatory γ -band (30–70 Hz) activity induced by a visual search task in humans. *Journal of Neuroscience*. 1997;17(2):722-34.