

fMRI Studies of Neural Mechanisms for Word
Processing in the Ventral Occipito-Temporal Cortex

March 2014

Geqi Qi

Graduate School of Natural Science and Technology

(Doctor's Course)

OKAYAMA UNIVERSITY

Abstract

Written words have been in use for only a few thousand years, which is not long enough for natural selection to produce a genetically specified brain region dedicated to visual word recognition. However, some previous brain imaging studies have suggested that a ventral occipito-temporal (vOT) region might be functionally specialized for visual word processing. To clarify the functional role of the vOT, two studies were conducted using functional magnetic resonance imaging (fMRI) techniques. In the first study, brain activations of literate and illiterate people were compared while they were judging whether two Chinese characters presented simultaneously on the screen were the same or not. The findings suggest that a direct relationship between visual word expertise and enhanced vOT response to written words should be taken with caution, since it can be affected by other factors such as task demand and attention. In the second study, effect of task demand and visual attention on vOT response to words was examined. The results suggest that the response property of the vOT is apt to change with task demand and level of attention, therefore underlining the importance to look more at the region's functional plasticity. To summarize, findings from the two studies suggest that the vOT is not functionally specialized for visual word processing, but is involved in visual processing of other objects as well.

Contents

CHAPTER 1 Introduction.....	1
1.1 Visual Word Recognition	2
1.2 Ventral occipito-temporal cortex	6
1.2.1 Specialization view	7
1.2.2 Generalization view	10
CHAPTER 2 Word Recognition in the Literate and Illiterate Brain.....	14
2.1 Background	15
2.2 Method	18
2.3 Results	21
2.4 Discussion	26
2.5 Conclusion	29
CHAPTER 3 Word and House Recognition in the vOT cortex	30
3.1 Background	31
3.2 Method	33
3.3 Results	38
3.4 Discussion	44

3.5 Conclusion	50
CHAPTER 4 General Discussion and Conclusion	51
4.1 General Discussion	52
4.2 General Conclusion	57
APPENDICES	60
I. Introduction to fMRI	60
II. BOLD Hemodynamic Response	63
III. Data Analysis with SPM	68
REFERENCES	78
PUBLICATIONS	85
ACKNOWLEDGEMENT	86

Chapter 1

Introduction

Summary

Reading is such a casual and automatic act in our daily life that we tend to forget what a remarkable skill it is and what convenience it has brought along to us. It is universally known that reading usually refers to converting the visual information of words into their corresponding sounds and meanings, and it can take years of training before we become fluent at it. We may have all at one point in our lives wondered the same question: how do we read? Advances in psychology and neuroscience over the last twenty years have shed light on the question. This chapter gives a brief introduction on the underlying mechanism of this feat.

1.1 Visual Word Recognition

Farmers are able to recognize their livestock among others in an instant manner just by observing their body shapes, hair colors and movements. Although the subtle differences between the animals can be very difficult for the rest of us to identify, the farmers are able to spot them with a minimum effort. Such superb skill of recognition is usually referred to as perceptual expertise (Garrigan & Kellman, 2008; Kellman & Garrigan, 2009). Similar skills are possessed by experts in a variety of disciplines ranging from dog-show judges to birdwatchers.

Although relatively few people specialize in the recognition of particular objects (e.g., birds, cars, dogs), it has been suggested that virtually all people who can read fluently are experts in the recognizing words (McCandliss et al., 2003a). What is the neural basis of such perceptual expertise? Noninvasive brain imaging techniques have enabled us to peek inside the brain during the emergence of perceptual expertise in word recognition (Grill-Spector & Malach, 2004). As humans acquire expertise there are dramatic changes in brain activity that complement the profound changes in processing speed and effort seen in behavioral data (Wandell et al., 2012). These changes involve learning, developing new representations, strategy shifts, and use of wider cues and approaches.

Like all other visual object recognition, written word recognition starts from our eyes. However, anatomy of our human eyes shows that only the center of the retina, which is called

the fovea, has resolution fine enough to let us recognize small prints. As shown in Figure 1.1A, the fovea is a small depression (approximately 1.5 mm in diameter) in the retina. This is the part of the retina in which high-resolution vision of fine detail is possible. The visual acuity decreases progressively away from fovea and towards the periphery (Figure 1.1B). As a result, we have to move our eyes constantly when reading to allow visual information of words to go through the particular part of the eye. This kind of scanning movement is often referred to as “saccades”.

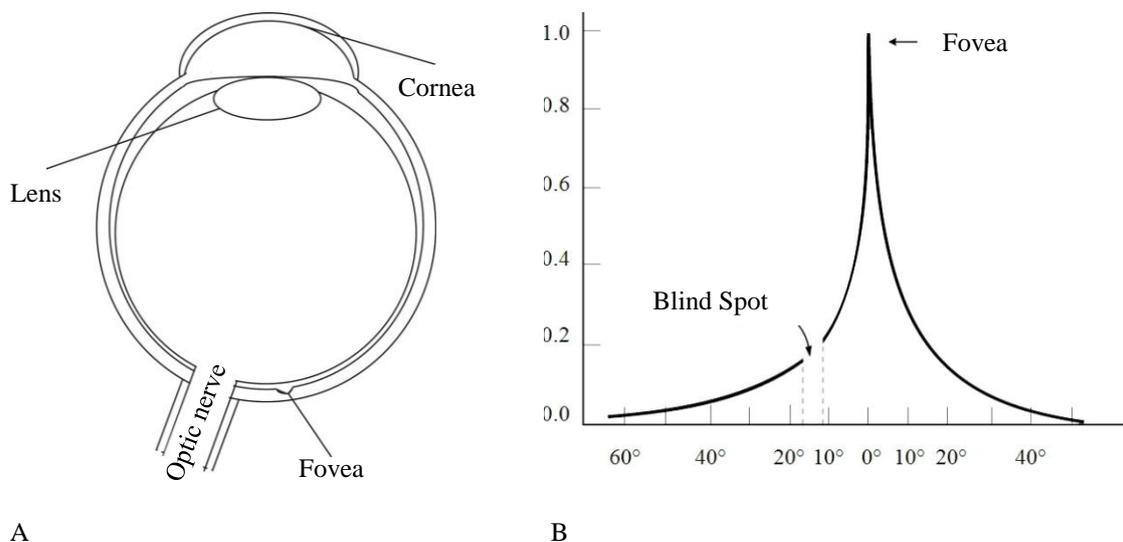


Figure 1.1. (A) Simple graphic illustrating the anatomy of the eye. Only fovea has resolution fine enough to let us recognize small prints. (B) The diagram shows the relative detail (vertical section) achieved by human eye in degrees from the fovea (horizontal section).

As you move your gaze from one word to another, the visual precision is shifted accordingly. Our eyes are able to accurately perceive only the precise point where the gaze is,

and the surroundings are blurred with haziness. As is shown in Figure 1.2, only four to five letters are seen with 100% acuity around our gaze. To compensate for this limitation, our brain adapts and prepares our gaze to always advance by 7-9 letters when we read. This finding may be counter intuitive to many since fluent readers are able to consume pages of writing with a speed of several hundred of words per minute.

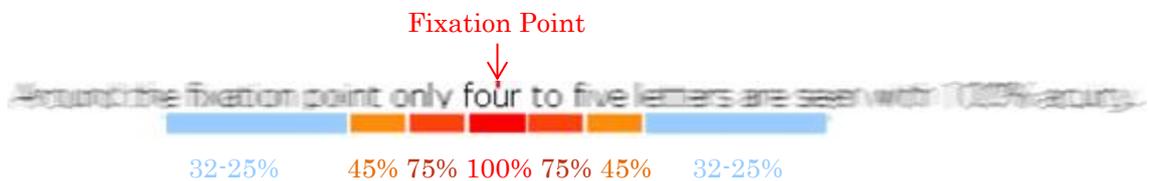


Figure 1.2. Reading is an intensive process in which the eye quickly moves to assimilate text. Very little is actually seen accurately. Around the fixation point only four to five letters are seen with 100% acuity.

Reading poses a difficult perceptual problem, which is to extract the invariance feature across words with various sizes and fonts (Figure 1.3). However, as we become more and more fluent at reading, word size can vary fiftyfold without impacting our reading speed. It is similar to the fact that objects and faces can be recognized regardless of the distance of two feet or thirty yards. Our ability to see invariance features demonstrates the tolerance of visual system to vast changes in scale. It also suggests that we have a letter normalization process in our visual system that resists transformation. Our capacity to recognize words does not depend on an analysis of their overall shape. Instead, it depends on a virtual dictionary, which is called mental

lexicon, to assign same address to same words regardless of their different shapes. Models of mental lexicon suggest that meaning is stored in our brain by decomposing a word into morphemes (smallest unit with meaning). Such decomposition is an essential step from vision to meaning (Turkish and Finnish morphemes can be assembled into large word with meaning equals a sentence). Evidences from priming effect studies showed that our visual system unconsciously snips out morphemes. For example, words which are visually different but share common morpheme can speed up each other's processing when presented sequentially (e.g. can & could). Also, words which have unrelated meaning but share common morpheme can prime each other (e.g. depart & department). Morphemes can be further decomposed into syllables (button → ba ton), and syllables into consonants and vowels (b a t t o n). This is another essential unit in reading to map grapheme to phoneme (tt → t). Overall, visual processing of words is like a tree, dissecting the words into morphemes then into syllables then into phonemes.

It had been proposed that there are two routes for reading: route one is a process to map vision to sound; and route two is a process to map vision to meaning. Evidence for route one comes from surface dyslexia. People with such reading disorder are able to read infrequent but regular words and unable to read frequent but irregular words. They have to first pronounce the words to understand the meaning. Evidence for route two came from brain injuries that

resemble deep/phonological dyslexia. People with such reading disorder are unable to pronounce regular but infrequent words. However, they can read out aloud irregular but frequent words or read a different word with related meaning. In children, both reading routes are poorly coordinated with each other, therefore depending largely on one route than the other. It takes years of practice to use both fluently for reading. Reading modals must be based on brain architecture, whose organization is characterized by multiple pathways. Therefore, two route modal is a useful approximation but a simplified version of the actual reading system.

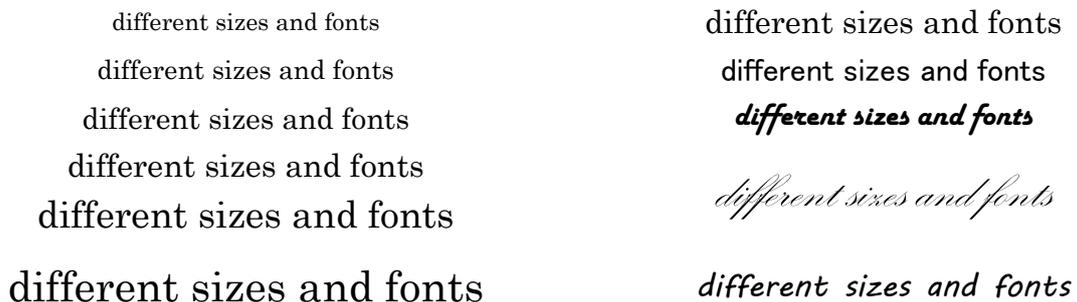


Figure 1.3. Visual invariance is one of the prime features of the human reading system. Our word recognition device meet two seemingly contradictory requirements: it neglects irrelevant variations in character shape, even if they are huge, but amplifies relevant differences, even if they are tiny.

1.2 Ventral Occipito-Temporal (vOT) Cortex

Previous imaging studies have localized a region in the ventral occipito-temporal (vOT) cortex that develops expertise for visual reading and extracts invariance features across various

shapes of words (Cohen et al., 2000; Op de Beeck et al., 2008). Developmental studies indicated that the vOT cortex becomes more and more sensitive to word stimuli while children learn to read (Wandell et al., 2012). Furthermore, vOT sensitivity to words increased with age and that the size of the increase correlated with the size of improvement in the ability to rapidly recognize a word from a nonword (Turkeltaub et al., 2003; Ben-Shachar et al., 2011). Since studies on the vOT have been dominantly visual reading, the region is sometimes referred to as the visual word form area (VWFA), which suggested a visual specific characteristic of this region (Cohen et al., 2000; McCandliss et al., 2003a). However, sequent studies questioned the specificity view and argued that the vOT is a stimulus general brain region, because many evidences showed that this region is strongly involved in processing of other complex visual stimuli such as pictures (Price & Devlin, 2003a). More profound objections came from studies of the blind literates (Buchel et al., 1998). For example, some studies found that comparable reading expertise can be achieved using Braille reading that in the blind people (Reich et al., 2011).

1.2.1 Specialization View

In 2000, a group of researchers led by Laurent Cohen and Stanislas Dehaene reported a case of word processing in the brain of patients suffering from left hemialexia following posterior callosal lesions (Cohen et al., 2000). Their study highlighted the damaged region

locating at the mid portion of left fusiform gyrus and hypothesized that this area is functionally specialized for processing of “visual word form” – a neural representation after primary visual processing and before phonological and semantic processing. This finding echoed the 1892 discovery of French neurologist Joseph-Jules Déjerine, who reported a selective impairment of word recognition in patient Mr. C after a stroke in the left visual brain system. Symptoms of the patients described above are very similar to those of people with pure alexia. More specifically, these patients have simply lost their ability to read while they could still normally speak, recognize faces and objects and even write. Normal people, on the other hand, have demonstrated consistent response properties to presented word stimuli in this region. For example, the region responds invariantly to words across a range of retinal positions, sizes and fonts. Laurent Cohen and colleagues, therefore, tentatively labeled this region of interest as the “visual word form area, VWFA”. One important proposal of VWFA hypothesis is that the left vOT is functionally specialized for visual word processing.

After years of learning to read, we are able to rapidly and effortlessly conjoin letters into a meaningful unit – word. The ability reflects an emergence of a perceptual expertise that facilitates us behaviorally regarding words, such as recognizing words rapidly from nonwords. It has been established in the past decade of neuroimaging studies that learning to read profoundly affects the anatomical and functional organization of the human brain (Pettersson et

al., 2001; Carreiras et al., 2009; Dehaene et al., 2010). One important change that occurs during the course of learning to read is the development of the ventral occipito-temporal (vOT) circuit that links visual and language systems (Price & Devlin, 2011; Wandell et al., 2012). Through years of reading, literate individuals establish efficient reciprocal projections between visual and language areas, which enables them to decode written words rapidly to acquire the sounds and meanings (McCandliss et al., 2003b). Such visual expertise is usually associated with increased vOT sensitivity to written words. For example, Ben-Shachar et al. (Ben-Shachar et al., 2011) measured children's (age 7–15 years) vOT responses to words and nonwords presented in different levels of noise. They found that the vOT sensitivity to words increased with age and that the size of the increase correlated with the size of improvement in the ability to rapidly recognize a word from a nonword. Consistent with this finding, the accuracy of literate adults in timed letter recognition was correlated with the level of vOT activation to written words (Garrett et al., 2000).

In literate adults, who are experts in visual word recognition, some response characteristics of the vOT suggest a relative insensitivity to variations in surface features of presented letters (i.e. letter case, font, size). For example, the vOT shows equally robust fMRI activations when words (versus control stimuli) are presented in a familiar format (e.g. 'table') and in a perceptually novel mixed-case format (e.g. 'tAbLe'). Furthermore, when a word is repeated

immediately, fMRI and N200 ERP responses associated with the vOT demonstrate sensitivity to word repetition that is independent of changes in letter case. Case invariance holds even for upper and lower-case letters that share no visual similarity but are related only by arbitrary cultural convention (e.g. a and A, g and G). This indicates that the abstract processes of the vOT go beyond size and shape normalization.

1.2.2 Generalization View

Soon after the VWFA hypothesis was proposed, it was questioned by other researchers. The VWFA is found to be activated during tasks that do not engage visual word form processing such as naming pictures, naming colors, repeating auditory words, reading Braille, and making manual action responses to pictures of meaningless objects (Price & Devlin, 2003b).

When subjects were asked to name, view, or verbally respond to pictures of objects, the VWFA was reported to be strongly activated (Murtha et al., 1999; Etard et al., 2000; Kherif et al., 2011). Furthermore, others have reported activation in this region for pictures of animals more than pictures of tools (Chao et al., 1999; Chao et al., 2002). Similarly, another study (van Turennout et al., 2000) has shown an effect of object learning in the VWFA, with less activation when the picture is familiar. Thus, pictures of common objects activate the same location of the VWFA and this activation can be enhanced by specific categories of object or reduced by priming. In addition, Mei and colleagues (Mei et al., 2010) revealed the important role of the

VWFA is not only in word processing but also in face processing and memory, therefore supported the view that the vOT cortex plays a general role in the successful processing and memory of different types of visual objects.

Some studies suggested that the vOT is crucial in some way for rapid decoding of complex visual stimuli—a process that reading is especially reliant upon (Devlin et al., 2006; Mycroft et al., 2009). Malach and colleagues (Levy et al., 2001; Hasson et al., 2002; Malach et al., 2002) proposed retinotopical organization in region which runs from the vOT to the collateral sulcus. Within this region, specific areas respond preferentially to different object categories (words, objects, houses, faces and animals). They (Malach et al., 2002) proposed that this graded functional separation reflects the visual demands of each type of stimuli and, in turn, variation of acuity across retinal eccentricity. Visual acuity, which refers to sensitivity to high spatial frequencies, is highest in the fovea and drops toward the parafoveal region (Fiset et al., 2006; Starrfelt et al., 2009). Foveal vision is delivered to the vOT and this region is preferentially active for categories of visual stimuli that require fine visual discrimination (e.g. faces and word)(Hasson et al., 2002). Other categories of visual stimuli (e.g. houses) activate areas medial to the vOT where parafoveal vision is primarily projected. In effect, there is a graded division of work across the vOT so that acuity-demanding visual categories dominate activation of the vOT, whereas categories which are less dependent on high-resolution foveal input can be processed

by peripherally biased regions on the basis of low spatial frequency information.

Additional evidences supporting generalization view on role of vOT come from its engagement in auditory word and Braille processing. A recent study (Striem-Amit et al., 2012) studied a group of congenitally fully blind adults trained to read through a visual-to-auditory sensory substitution device (SSD). The result demonstrated that the vOT is selective to letters over all other tested stimuli, including the SSD transforms of both low-level textures and visually complex objects, regard-less of sensory modality, visual experience, long-term familiarity, or expertise with the script. These findings suggest a strong feature tolerance in this area, which generalizes even beyond sensory input modality and early sensory experience, while maintaining the relative category selectivity to words. Another previous study (Yoncheva et al., 2010) contrasted two equally challenging active listening tasks, performed on the same complex auditory stimuli (words overlaid with a series of 3 tones).The results showed that selective attention to speech, relative to attention to melody, was associated with BOLD increases during fMRI in the vOT. Thus, selective attention to speech can topographically tune extrastriate cortex, leading to increased activity in vOT relative to surrounding regions, in line with the well-established connectivity between areas related to spoken and visual word perception in skilled readers. By isolating the impact of top-down selective auditory attention to phonological information, this and similar studies may prove valuable for future investigations into how

attention to phonology influences reading acquisition and the rise of selective activation of the vOT.

Recent evidences suggest that blind people use areas of the cerebral cortex normally reserved for vision during Braille reading and other nonvisual tasks involving tactile discrimination (Burton et al., 2002). An outstanding example of such areas is the vOT cortex which has been suggested to be specific to visual word form processing. Early in the 1998, Buchel et al. (Buchel et al., 1998) tested whether the visual component of reading is essential for activation of vOT cortex by comparing cortical activations elicited by word processing in congenitally blind, late-blind and sighted subjects using functional neuroimaging. Despite the different modalities used (visual and tactile), all groups of subjects showed a common activation of vOT cortex by words relative to non-word letter-strings. Reich et al. (Reich et al., 2011) observed that activation during Braille reading in blind individuals peaks in the VWFA, with striking anatomical consistency within and between blind and sighted. Furthermore, the VWFA is reading selective when contrasted to high-level language and low-level sensory controls.

Chapter 2

Word Recognition in the Literate and Illiterate Brain

Summary

Visual word expertise is typically associated with enhanced ventral occipito-temporal (vOT) cortex activation in response to written words. Dehaene et al. (2007) utilized a passive viewing task and found that vOT response to written words was significantly stronger in literate compared to the illiterate subjects. However, recent neuroimaging findings have suggested that vOT response properties are highly dependent upon the task demand. Thus, it is unknown whether literate adults would show stronger vOT response to written words compared to illiterate adults during other cognitive tasks, such as perceptual matching. This issue was addressed by comparing vOT activations between literate and illiterate adults during a Chinese character and simple figure matching task. Unlike passive viewing, a perceptual matching task requires active shape comparison, therefore minimizing automatic word processing bias. Although the literate group performed better at Chinese character matching task, the two subject groups showed similar strong vOT responses during this task. Overall, the findings indicate that the vOT response to written words is not affected by expertise during a perceptual matching task, suggesting that the association between visual word expertise and vOT response may depend on the task demand.

2.1 Background

It has been established in the past decade of neuroimaging studies that learning to read profoundly affects the anatomical and functional organization of the human brain (Petersson et al., 2001; Carreiras et al., 2009; Dehaene et al., 2010). One important change that occurs during the course of learning to read is the development of the ventral occipito-temporal (vOT) circuit that links visual and language systems (Price & Devlin, 2011; Wandell et al., 2012). Through years of reading, literate individuals establish efficient reciprocal projections between visual and language areas, which enables them to decode written words rapidly to acquire the sounds and meanings (McCandliss et al., 2003). Such visual expertise is usually associated with increased vOT sensitivity to written words. For example, Ben-Shachar et al. (Ben-Shachar et al., 2011) measured children's (age 7–15 years) vOT responses to words and nonwords presented in different levels of noise. They found that the vOT sensitivity to words increased with age and that the size of the increase correlated with the size of improvement in the ability to rapidly recognize a word from a nonword. Consistent with this finding, the accuracy of literate adults in timed letter recognition was correlated with the level of vOT activation to written words (Garrett et al., 2000).

These findings gave rise to the hypothesis that extensive experience with written words enhances the vOT activation to this particular domain of visual stimuli (Cohen et al., 2000;

McCandliss et al., 2003; Baker et al., 2007). However, significant evidence suggests that additional factors are involved (Price & Devlin, 2003, 2011). For example, Song et al. (Song et al., 2010) used a set of novel objects to train two groups of subjects in two different task contexts (association and discrimination tasks). After the training, increased vOT activations for the trained objects were observed only in the subjects who learned the objects under the association task, but not in the group who learned under the discrimination task. This finding demonstrated that the effects of visual expertise on vOT activation depend on how the expertise is acquired during learning to encode a domain of stimuli. Many studies have compared the vOT activation to written words with activation to other types of stimuli. Although early findings indicated preferred vOT activation to written words compared to other objects (McCandliss et al., 2003; Baker et al., 2007), an increasing number of studies have shown that vOT activation changes with the task, even when the stimulus, attention and response times are controlled (Yoncheva et al., 2010; Twomey et al., 2011). For example, Joseph et al. (Joseph et al., 2006) compared vOT activation to written words and pictures in literate adults and found that the vOT cortex showed higher activation for words than for pictures during a passive viewing task, but not during a perceptual matching task. These findings further demonstrated that the vOT response to written words is highly dependent on the task demands.

Given these premises, one could assume that literate adults would show higher vOT

activation to written words compared to illiterate adults. However, this enhanced activation should be observed only under certain circumstances. Only one study has compared the vOT response to written words between literate and illiterate adults (Dehaene et al., 2010). In that study, Dehaene et al. utilized a passive viewing task. They found that the vOT cortex was significantly more activated in literate subjects than in illiterate subjects. However, since expertise with written words automatizes word processing in vOT, an automatic processing bias may occur during passive viewing task (Gauthier, 2000; McCandliss et al., 2003). No study has compared the vOT response to written words between literate and illiterate adults in other cognitive tasks, such as perceptual matching. Unlike passive viewing, a perceptual matching task emphasizes visuospatial processing and requires active shape comparison of the paired stimuli, thereby minimizing the automatic word processing bias (Wojciulik et al., 1998). In previous study (Wu et al., 2012), a perceptual matching task was applied and activations between the literate and illiterate groups were directly compared on a whole brain scale. It was found that the literates showed significantly stronger response in the bilateral inferior parietal lobule (IPL) during Chinese character matching task. However, processing in low-level visual areas of literate and illiterate brains was not explicitly compared. Therefore, it is still unclear how vOT might be influenced by literacy under a perceptual matching task. In this paper, the issue was addressed by comparing vOT activations between literate and illiterate adults during

Chinese character and simple figure matching tasks.

2.2 Method

2.2.1 Subjects

Twenty-four right-handed subjects were included in this study. All subjects were healthy native Chinese speakers from Northeast China. Half of the subjects were literate (mean age, 47 years; range, 39–55 years) while the other half were illiterate (mean age, 48 years; range, 34–57 years). The illiterate subjects had never entered school for social reasons and had no knowledge of reading or writing. The literate and illiterate subjects had a normal intelligence quotient; the social-economy index of the illiterate subjects was significantly lower than that of the literate subjects at the time the literate subjects entered school (Table 2.1). None of the subjects were on medication, nor did they have any history of neurological or psychiatric disorders. All subjects provided written informed consent prior to participation in the experiment. The present study was approved by the Ethics Committees of the Shengjing Hospital of the China Medical University.

2.2.2 Experimental Design

The Chinese characters were selected from fifth grade primary school textbooks. Then recognition and reading tests were conducted with the subjects. Two simple figures (a circle and a square) were selected as the non-language stimuli. Characters or figures were combined into

pairs; half of the pairs included either two identical figures or two identical characters. Figure pairs consisted of two figures, either circles, squares or both, yielding four possible pairs. For the pairs of different stimuli, the characters were made to look similar, and parts of the sub-character units were the same; for example, 衫 and 杉 have the same sub-character (彡). In total, 36 pairs of characters were used in this experiment, with 36 trials delivered for each experimental condition.

Table 2.1. Questionnaire results from the literate and illiterate subjects

	Literate	Illiterate	Sig.
Age	45.58	48.33	P=0.28
IQ Score	110.17	90.25	P<0.0001
Social-economy	458.33	102.08	P<0.0001

A random-interval event design was adopted; the characters and figures were randomized over the session. Each pair of characters or figures was presented through a projector for 4000 ms, with a random inter-trial interval of 2000, 4000 or 6000 ms between the stimuli. The subjects were asked to fixate on a small crosshair during the task. During each trial, the subjects were shown a pair of stimuli (Chinese characters or simple figures) and asked to judge whether the two stimuli were the same as quickly as possible. The stimuli were presented using E-Prime software (Psychology Software Tools, INC., Sharpsburg, PA, USA).

2.2.3 fMRI Data Acquisition

Functional imaging data were acquired using a 3 T Philip signal scanner at the Shengjing Hospital of China Medical University. A total of 289 fMRI volume images were collected during each session. Echo-planar images (EPIs) sensitive to the blood oxygen level-dependent (BOLD) response of 36 slices were acquired using the following scan parameters: TR = 2000ms, matrix size = 64×64, voxel size = 3×3×3.5 mm³, slice thickness = 3.5 mm, slice gap = 0 mm, effective echo time (TE) = 30 ms, flip angle (FA) = 90°. After functional scanning, three-dimensional (3D) T1-weighted anatomical images with a 1 mm³ isotropic voxel size (matrix 256×256×182) were acquired for cortex normalization and cortex surface reconstruction.

2.2.4 fMRI Data Analysis

The data were processed using the SPM5 software package (Wellcome Trust Centre for Neuroimaging, UCL, UK). The first three scans were discarded from the analysis to eliminate nonequilibrium effects of magnetization. The scans were realigned, normalized, smoothed (8-mm, Gaussian spatial filter), and filtered (high-pass filter set at 128 s). The resulting images had cubic voxels of 2×2×2 mm³. Individual subjects' activation t-maps were generated using a general linear model in which the time series was convolved with the canonical hemodynamic response function (HRF) to model the BOLD response. A group-level random effects analysis

was performed by conducting a one-sample t-test across all individual subjects in each group, and a two-sample t-test to directly compare the whole brain activations between the two groups. Main effects were examined by contrasting each type of stimuli with the fixation stimulus. Both of the contrasts used a corrected voxelwise threshold of $P < 0.05$, with a minimum volume size of 100 mm³.

MarsBaR software ("Matthew Brett, Jean-Luc Anton, Romain Valabregue, Jean-Baptiste Poline. Region of interest analysis using an SPM toolbox [abstract] Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. Available on CD-ROM in NeuroImage, Vol 16, No 2,") was used to calculate the percent BOLD signal change in the left and right ROI in the vOT cortex. The left vOT ROI was defined as a sphere of all the voxels within a 6 mm radius surrounding the average of the visual word form area (VWFA) peaks identified in previous studies (Bolger et al., 2005). The right vOT ROI was defined as the mirror-reversed sphere in the right hemisphere. The percent signal changes were then compared using a repeated-measures ANOVA in SPSS 13.0 (SPSS Inc., Tokyo, IL).

2.3 Results

2.3.1 Behavioral Data

The response times and accuracy of each subject during the character and figure tasks are

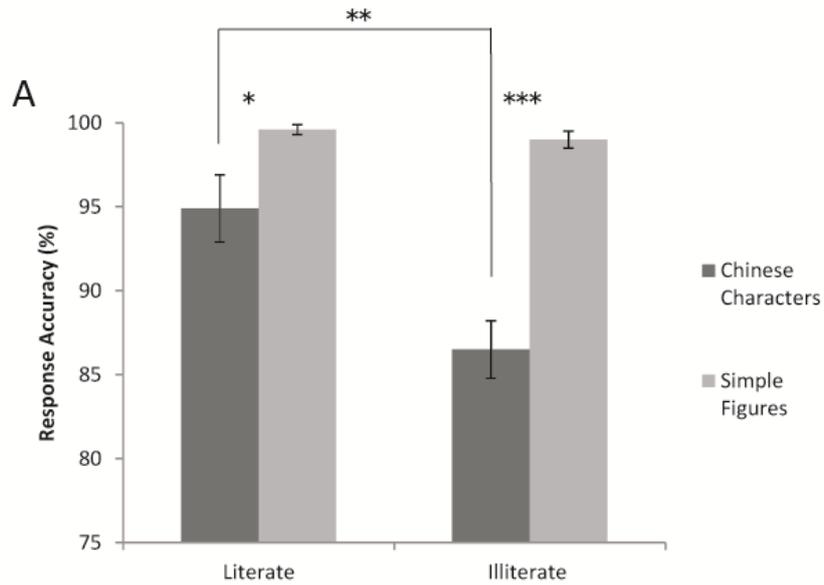


Figure 2.1. (A) Response time of the two tasks. Two-way repeated-measures ANOVA demonstrated a significant main effect of task ($P < 0.0001$) and group ($P = 0.05$), as well as a significant task \times group interaction ($P = 0.002$).

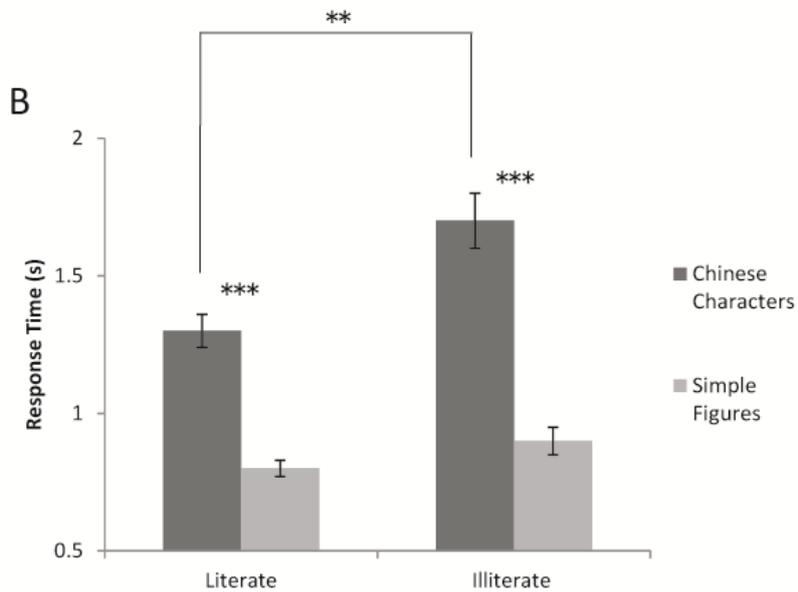


Figure 2.1. (B) Response accuracy of the two tasks. Two-way repeated-measures ANOVA demonstrated a significant main effect of task ($P < 0.0001$) and group ($P = 0.013$), as well as a significant task \times group interaction ($P = 0.01$). Pairwise comparisons with Bonferroni corrections showed significant differences between the two groups and the two tasks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$, error bars, Mean \pm SEM).

shown in Figure 2.1 The results show longer response times and lower accuracies for characters compared to figures. For response times, a two-way repeated-measures ANOVA demonstrated a significant main effect of the task [$F(1, 22) = 205.94, P < 0.0001$], a main effect of the group [$F(1, 22) = 9.48, P = 0.05$], and a significant task \times group interaction [$F(1, 22) = 11.75, P = 0.002$].

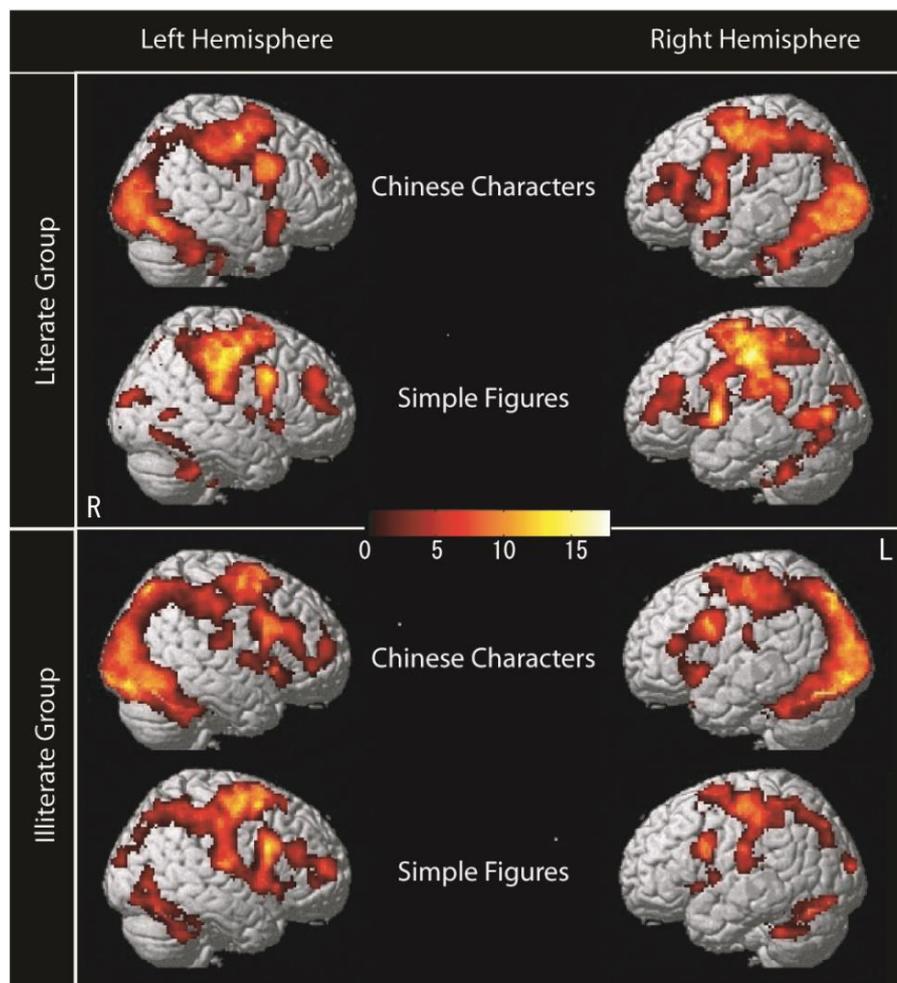


Figure 2.2. Whole brain activations of the two subject groups for Chinese characters and simple figures in contrast to the rest condition. The activation intensity is indicated by regions in red to yellow. A voxel threshold of $P < 0.05$, corrected with FDR, and a minimum volume size of

100 mm³ were used.

2.3.2 Functional Brain Activity

As shown in Figure 2.2, the two subject groups showed strong bilateral activation for both Chinese character and simple figure matching tasks. The Chinese characters yielded stronger activation in the vOT cortex than the simple figures in both literate and illiterate subjects (Figure 2.3). The whole brain activations between the groups were also directly compared using two-sample t-test, and found that the literates showed significantly stronger response in the bilateral inferior parietal lobule (IPL) during Chinese character matching task (Wu et al., 2012).

To further investigate the vOT activation, the percent BOLD signal changes were calculated in the left and right vOT masks for the two perceptual matching tasks (Figure 2.4). A two-way repeated-measures ANOVA demonstrated a significant main effect of the task in the left [$F(1,22)=117.90$, $p<0.0001$] and right [$F(1,22)=125.14$, $p<0.0001$] vOT cortex, with no significant effect of group or task \times group interaction. Pairwise comparisons with Bonferroni corrections showed a significant difference between the two tasks in the left vOT cortex (literate: [$F(1,22)= 57.47$, $p<0.0001$], illiterate: [$F(1,22)= 60.46$, $p<0.0001$]) and in the right vOT cortex (literate: [$F(1,22)= 52.73$, $p<0.0001$], illiterate: [$F(1,22)= 87.99$, $p<0.0001$]).

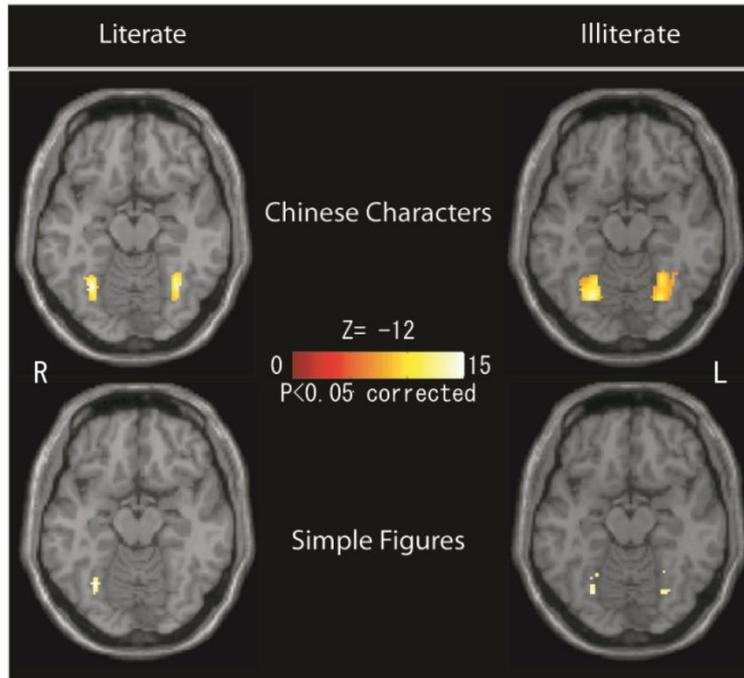


Figure 2.3. Activations revealed by the vOT mask under the contrast of Chinese characters and simple figures vs. rest. The left column shows the vOT activation in literate subjects, while the right column shows the vOT activation in illiterate subjects. The activation intensity is indicated by regions in red to yellow. A voxel threshold of $P < 0.05$, corrected with FDR, and a minimum volume size of 0 mm^3 were used.

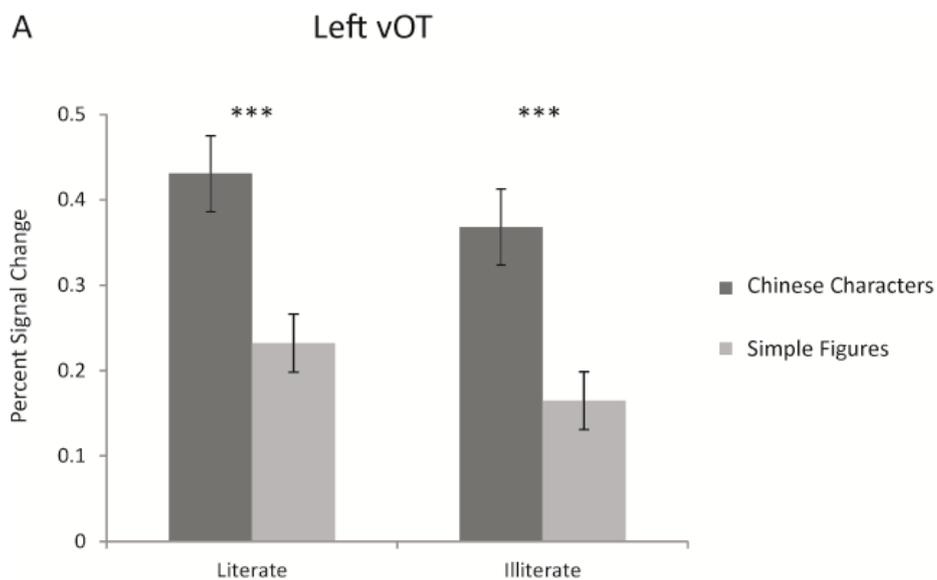


Figure 2.4. (A) Percentage BOLD signal changes in the left vOT cortex. A two-way

repeated-measures ANOVA demonstrated a significant main effect of task in the left vOT cortex ($P < 0.0001$, respectively). The activation levels for Chinese characters and simple figures were significantly different in the left vOT (literate, $p < 0.0001$; illiterate, $p < 0.0001$). $***p < 0.0001$, error bars, Mean \pm SEM.

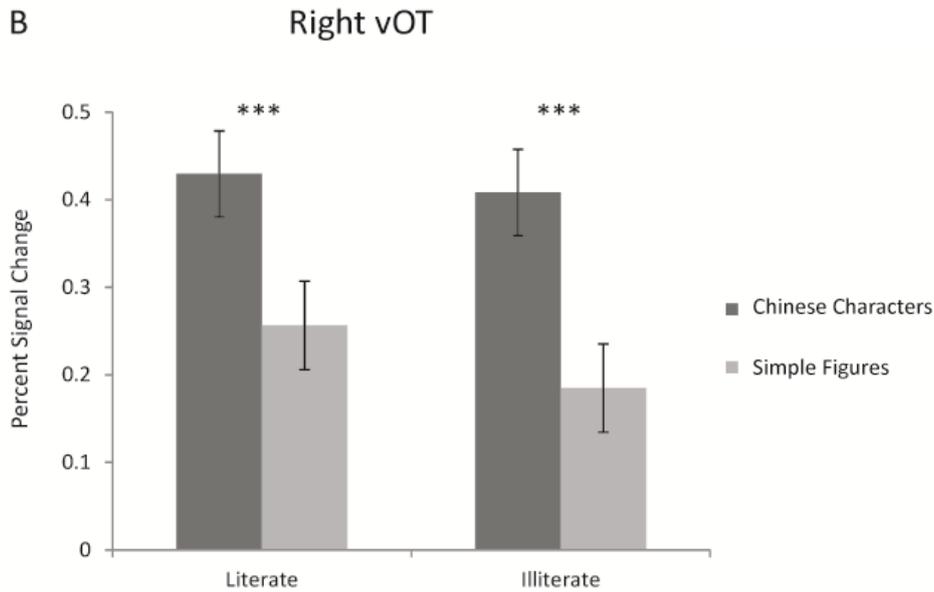


Figure 2.4. (B) Percentage BOLD signal changes in the right vOT cortex. A two-way repeated-measures ANOVA demonstrated a significant main effect of task in the right vOT cortex ($P < 0.0001$, respectively). The activation levels for Chinese characters and simple figures were significantly different in the right vOT (literate, $p < 0.0001$; illiterate, $p = < 0.0001$). $***p < 0.0001$, error bars, Mean \pm SEM.

2.4 Discussion

In the present study, whether literate adults would show enhanced vOT response to written words compared to illiterate adults during a perceptual matching task was examined. We compared the vOT activations between literate and illiterate subjects during Chinese character and simple figure matching tasks. As a result, we found a significant difference between the two

tasks, but not between the two subject groups. These findings suggest that the vOT response to written words is not affected by visual word expertise during a perceptual matching task.

As shown in Figure 2.1, the literate group had a significantly better performance on the Chinese character matching task, whereas the performances on the simple figure matching task were equal between the two groups. This result is consistent with previous findings suggesting extensive experience with certain object domains leads to visual expertise that allows the visual system to process the objects efficiently (McCandliss et al., 2003). Such fast processing was observed across a wide domain of object categories, including faces, cars, birds, greebles and words (Gauthier et al., 2000; Tarr & Gauthier, 2000; McCandliss et al., 2003). When we are learning to encode written words, we associate the visual shapes with sounds and meanings. The necessity of linking vision to language requires computations executed in the vOT cortex (Price & Devlin, 2011; Wandell et al., 2012). As this experience accumulated and expertise was acquired, the encoding process in vOT became automatized (Gauthier, 2000). For example, a study combining fMRI and ERP methods showed that the vOT produced measurable responses to words, even under conditions that prevent awareness of the visual presentation of a word (Dehaene et al., 2001). These evidences are consistent with the result of Dehaene et al. (Dehaene et al., 2010) that the vOT cortex in literate subjects was significantly more activated than in illiterate subjects during a passive viewing task. With minor cognitive demand, the

passive viewing task could make way for automatic processing in the expert vOT, thus causing stronger activation in literate subjects than in illiterate subjects.

In the present study, a matching task that emphasized visuospatial processing was employed. Unlike passive viewing, a perceptual matching task requires detailed shape comparison of the paired stimuli. Thus, the vOT response to Chinese characters would not be produced in a strictly automatic stimulus-driven manner during a perceptual matching task (Wojciulik et al., 1998). Concordantly, the results of the present study showed that vOT activation levels did not vary between the literate and illiterate subject groups (Figs. 3, 4). It may suggest that the vOT in both subject groups was involved in a general visuospatial processing (Price & Devlin, 2003; Reinke et al., 2008; Price & Devlin, 2011). And because the functional load imposed by complex stimuli is higher than that imposed by simple stimuli (Marcar et al., 2004), the Chinese characters increased the metabolism in vOT more compared to the simple figures (Figs. 3, 4). Strong vOT response to Chinese characters in illiterate subjects may also due to an extra attentional modulation, as they tend to require more visual attention than the literate subjects during visuospatial processing (Li et al., 2006; Petersson et al., 2007). Although visual word expertise was largely confined to the vOT cortex, several previous neuroimaging studies have suggested that expertise effects may be expressed across the entire cortex, reflecting a wider cortical network (Gauthier et al., 2000; Xu, 2005; Harel et al., 2010).

Indeed, previously reported results of a whole brain analysis (Wu et al., 2012) showed that compared to the illiterate subjects, the literate subjects had significantly higher activation in the bilateral IPL during the Chinese character matching task. The IPL has been implicated in language processing (Jung-Beeman, 2005) and literate people tend to have greater gray matter density in this region compared to illiterate people (Carreiras et al., 2009). Overall, these results suggest that the association between visual word expertise and vOT response may depend on the task demand.

2.5 Conclusion

In summary, literate and illiterate adults had similarly strong vOT responses to Chinese characters during perceptual matching task, although literate adults performed better at this task. Overall, the findings indicate that the vOT response to written words is not affected by visual word expertise during a perceptual matching task and suggest that the association between visual word expertise and vOT response may depend on the task demand.

Chapter 3

Word and House Recognition in the vOT cortex

Summary

Many studies have reported that blood oxygen level–dependent (BOLD) activity in object-selective extrastriate areas is enhanced by visual attention. However, only a few studies have compared the effects of attention between these areas. In this study, the influence of small and big noise level on BOLD responses was examined in the word-selective region (visual word form area, VWFA) and house-selective region (parahippocampal place area, PPA) during Chinese character and house pictograph discrimination task. The results showed that the BOLD response in the VWFA and PPA are modulated differently by visual attention. More specifically, the selective response in the VWFA is more easily affected by attentional demand than that in the PPA. These results may suggest that visual attention is distributed differently along the extrastriate cortex.

3.1 Background

Extrastriate visual areas are known to exhibit category-selective activations, responding much more strongly to their preferred stimuli than to other stimuli (Grill-Spector & Malach, 2004; Op de Beeck et al., 2008). Most robust category-selectivity have been observed in cortical regions that respond preferentially to images of objects (lateral occipital complex, LOC), to faces (the fusiform face area, FFA)(Kanwisher et al., 1997), to houses (the parahippocampal palce are, PPA) (R. Epstein & Kanwisher, 1998) and to written words (the visual word form area, VWFA) (Cohen et al., 2000). Such selective activations can be invariant to changes in the appearance of objects caused by viewing conditions such as colors and luminance contrasts (McCandliss et al., 2003; Grill-Spector & Malach, 2004; Rolls & Stringer, 2006). These regions lie in the ventral visual pathway anterior to early retinotopic cortex. The FFA and the VWFA have been described in a stretch of cortex bounded by the fusiform gyrus laterally, with the former locating anterior to the latter region. The PPA, on the other hand, is situated on the parahippocampal gyrus medially.

Although category-selective activation in the extrastriate visual areas exhibits perceptual invariance, it can also be modulated by attentional and task demands. For example, a recent study (Xu et al., 2012) adopting a new approach of Multi-Voxel Pattern Analysis demonstrated that the within-condition classification accuracies in the object-selective regions were critically

dependent on attention levels. Task and attentional modulation has been observed across the extrastriate visual areas, including the FFA(Wojciulik et al., 1998), PPA(Wendelken et al., 2011) and VWFA(Starrfelt & Gerlach, 2007; Song et al., 2010). In addition, a few previous evidences reported that the effects of attention were different between extrastriate areas. For example, Tallon-Baudry et al. (Tallon-Baudry et al., 2005) showed that attentional modulations of gamma oscillations in lateral occipital cortex and the fusiform gyrus were very different both in time-course and direction of modulation.

Some researchers suggested that different task and attentional modulation reflects distinct functions of these extrastriate visual areas. Levy et al. (Levy et al., 2001; Hasson et al., 2002) proposed a functional criteria which differentiates the object-selective regions. According to their view, the distributed locations of these regions project onto an eccentricity map, with the FFA and VWFA manifesting a foveal bias (big spatial resolutions), and the PPA a peripheral field bias (small spatial resolutions). Consistently, Reddy et al. reported different attentional modulation of blood oxygen level–dependent (BOLD) response in the FFA and PPA under same experimental condition(Reddy et al., 2007). However, no study has compared attentional modulations between the VWFA and PPA. The VWFA develops its functional selectivity much later in life compared to other object-selective regions. While it prefers central representation as the FFA, it is also very sensitive to high spatial frequency visual information as the PPA.

To investigate and compare task/attentional modulation in the VWFA and PPA, the influence of small (revealing 95% sensory evidence) and big (revealing 75% sensory evidence) noise levels on the BOLD response was examined in these two regions during a categorical discrimination task. House pictographs and Chinese characters were used as visual stimuli that shared similar big spatial frequency square configurations and a uniform gray color.

3.2 Method

3.2.1 Subjects

Thirteen healthy volunteers participated in the fMRI experiment (all male with a mean age of 23.4 years) and eight healthy volunteers in the behavioral experiment for psychophysical calibration (all male with a mean age of 22.8 years; five of them also participated in the fMRI experiment). All were right-handed, had normal or corrected-to-normal vision, and had no past neurological conditions, psychiatric history or structural brain abnormality. Informed consent was obtained according to procedures approved by the Okayama University ethics committee. Visual stimuli (approximately $5^\circ \times 5^\circ$) were projected onto a screen located outside the scanner at an apparent distance of approximately 60 cm when viewed through a mirror immediately in front of the subjects' eyes.

3.2.2 Psychophysical calibration

Sixty Chinese characters (2008, Kanji frequency table, Agency for Cultural Affairs, Japan)

(Matsuo et al., 2010) and sixty house pictographs were used. The Chinese characters and house pictographs were 240 x 240 pixel gray-scale images, superimposed by variable amounts of white noise in squares of 8 x 8 pixels. During the psychophysical calibration session, a Chinese character or a house pictograph appeared centrally for 1 s every 3 s. The subjects responded by pressing a key to judge whether the stimulus presented was a word/house or not. As shown in Figure 3.1.A, 60 trials for each stimulus category (character and house) with 6 equally spaced noise levels (0%, 15%, 30%, 45%, 60% and 75%) were randomly presented. A probit analysis of binomial responses based on maximum likelihood estimation provided the threshold and slope of the psychometric function, describing the probability of giving a word/house response as a function of the noise level in the stimuli. Interpolation of the individual psychometric function allowed us to select two noise levels for each individual, yielding 75% (a big noise level) and 95% (a small noise level) word and house responses (Figure 3.1.B).

3.2.3 Localizer scan

To identify VWFA and PPA, a localizer scan was conducted for each subject before the main experiment. Participants performed a one-back memory task during the localizer scan. They viewed sixteen alternating blocks (20 s) containing intact images taken from four different categories: houses, Chinese characters, textures or faces. Each image was presented for 800 ms and was interleaved with fixation periods of 200 ms. Each condition was repeated four times

and arranged in a counterbalanced block design.

3.2.4 Main experiment

Images of Chinese characters and house pictographs with two evidence/noise levels were presented using the Presentation software (<http://www.neurobs.com>). Stimuli (Figure 3.1.A) from the four conditions (Chinese character with small noise level, Chinese character with big noise level, house pictograph with small noise level, house pictograph with big noise level) were presented randomly for 1 s (Figure 3.1.C). After a delay (1-5 s) and a response cue (300 ms), subjects reported whether they had seen a word or a house by pressing their index or middle finger button, respectively.

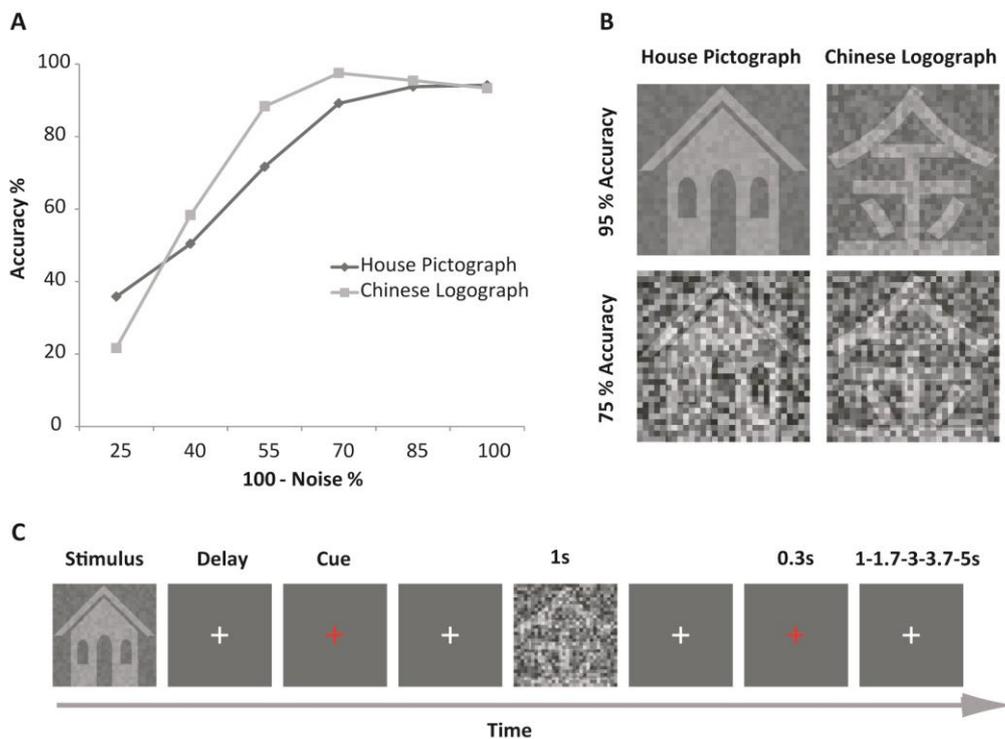


Figure 3.1. Experimental task. Subjects decided whether an image presented on a screen was a word or a house. (A) Result of behavioral study to assess the amount of noise to be added to

the stimuli. (B) In the fMRI experiment, Chinese characters and house pictographs that were either easy (95% sensory evidence, low level of noise, A top) or difficult (82% sensory evidence, high level of noise, A bottom) were used to discriminate. (C) Event-related fMRI design. Stimuli were presented for 1 s, and the subjects responded with a button press after a forced delay (response cue shown for 300 ms, delay 1–5 s).

3.2.5 fMRI Data Acquisition

Functional imaging data were acquired using a 3 T Siemens signal scanner at the Okayama University Hospital. A total of 220 fMRI volume images were collected during each session. Echo-planar images (EPIs) sensitive to the blood oxygen level-dependent (BOLD) response of 32 slices were acquired using the following scan parameters: TR = 2 s, matrix size = 128×128, voxel size = 3×3×3 mm³, slice thickness = 3 mm, slice gap = 0 mm, effective echo time (TE) = 30 ms, and flip angle (FA) = 90°. After functional scanning, three-dimensional (3D) T1-weighted anatomical images with a 1 mm³ isotropic voxel size (matrix 256×256×182) were acquired for anatomical comparison.

3.2.6 Data Analysis

The data were processed using the SPM8 software package (Wellcome Trust Centre for Neuroimaging, UCL, UK). The first three scans were discarded from the analysis to eliminate nonequilibrium effects of magnetization. The scans were realigned, normalized, smoothed (8-mm, Gaussian spatial filter), and filtered (high-pass filter set at 128 s). The resulting images had cubic voxels of 2×2×2 mm³. Individual subjects' activation t-maps were generated using a

general linear model in which the time series was convolved with the canonical hemodynamic response function (HRF) to model the BOLD response. A group-level random effects analysis was performed by conducting a one-sample t-test across all individual subjects. The main effects were examined by contrasting each condition of stimulus with the fixation stimulus. The localizer scan was used to define VWFA voxels and PPA voxels in each individual, and all further analyses were performed across the significantly activated voxels. The contrasts used an uncorrected threshold of $P < 0.05$ to allow subjects with weak selective voxels to also contribute to the ROI analyses. MarsBaR software was used to calculate the percent BOLD signal change in the VWFA and PPA. Table 3.1 shows the individual activation peaks in the VWFA and PPA. A total of 11 out of the 13 subjects showed significant activation in the VWFA, and 2 among the 11 subjects also showed significant activation in the right VWFA. All of the 13 subjects showed significant activation in the right PPA, as did 9 subjects in the left PPA. The percent signal changes were then compared between the experimental conditions using repeated-measures ANOVA in SPSS 13.0 (SPSS Inc., Chicago, IL).

Table 3.1. The individual activation peaks in the VWFA and PPA. A total of 11 out of the 13 subjects showed significant activation in the VWFA, and 2 among the 11 subjects also showed significant activation in the right VWFA. All of the 13 subjects showed significant activation in the right PPA, as did 9 subjects in the left PPA.

Subjects	(Visual Word Form Area) VWFA						(Parahippocampal Place Area) PPA					
	Left hemisphere			Right hemisphere			Left hemisphere			Right hemisphere		
	x	y	z	x	y	z	x	y	z	x	y	z
Sub1	-44	-52	-22							26	-18	-32
Sub2	-46	-48	-18				-28	-42	-14	24	-42	-14
Sub3	-42	-56	-24	50	-54	-24	-26	-42	-10	34	-36	-14
Sub4	-46	-44	-16				-30	-44	-10	34	-40	-14
Sub5	-44	-50	-20				-24	-32	-22	26	-44	-14
Sub6							-24	-38	-18	26	-38	-18
Sub7	-50	-48	-22	30	-82	-22				28	-56	-10
Sub8							-26	-54	-10	24	-60	-10
Sub9	-42	-54	-16				-30	-32	-22	38	-36	-16
Sub10	-52	-46	-20							34	-50	-8
Sub11	-44	-46	-16				-24	-48	-14	28	-46	-14
Sub12	-38	-56	-18							14	-52	2
Sub13	-42	-48	-16				-28	-36	-18	32	-44	-14

3.3 Results

The (BOLD) signal time series was recorded with functional magnetic resonance imaging (fMRI) during a categorization task using an event-related fMRI design (Figure 3.1.C). During a typical trial, either a Chinese character or a house pictograph was centrally presented, and the subjects reported whether they had seen a word or a house by pressing two buttons after a delay and a cue. The amount of noise added to the visual stimuli was assessed in a psychophysical calibration session prior to the fMRI experiment. The responses of the subjects were fitted to a

psychometric function, which was then interpolated to select a small level of noise (95% sensory evidence) and a big level of noise (75% sensory evidence) in the visual stimuli (Figure 3.1.B).

3.3.1 Main Effects for the 4 Experimental Conditions

A repeated measures two-way ANOVA was conducted to assess the behavioral result for the main experiment. As shown in Figure 3.2, the analysis demonstrated a main effect of the noise levels [$F(1, 12) = 44.837, P < 0.0001$], and no significant effects of the categories and interactions between the noise level and the category were found. Pairwise comparisons with Bonferroni corrections showed that a significant difference existed between the activation levels for Chinese characters with small and big level of noise [$F(1, 12) = 21.08, P < 0.001$]. The activation difference between the house pictographs with small and big level of noise was also significant [$F(1, 12) = 10.356, P < 0.007$].

As shown in Figure 3.3, extensive extrastriate activation was found for all of the 4 experimental conditions compared to fixation. Significantly activated areas included bilateral fusiform gyrus, lingual gyrus, calcarine sulcus, inferior occipital gyrus, inferior temporal gyrus, and left middle occipital gyrus. The bilateral parahippocampal gyrus activation was also significant to the house pictograph with both small and big levels of noise and to the Chinese characters with big level of noise.

3.3.2 Percent Signal Changes in the VWFA and PPA

Using the localizer scan, voxels in the VWFA that responded more to Chinese characters than to houses and voxels in the PPA that responded more to houses than to Chinese characters were identified in each subject. The word-selective activation was lateralized to the VWFA (11 out of 13 subjects), while house-selective activation included voxels from both hemispheres (9 out of 13 subjects). To further assess the roles of the VWFA and PPA during the categorization task, the relative responses to the four experimental conditions were compared.

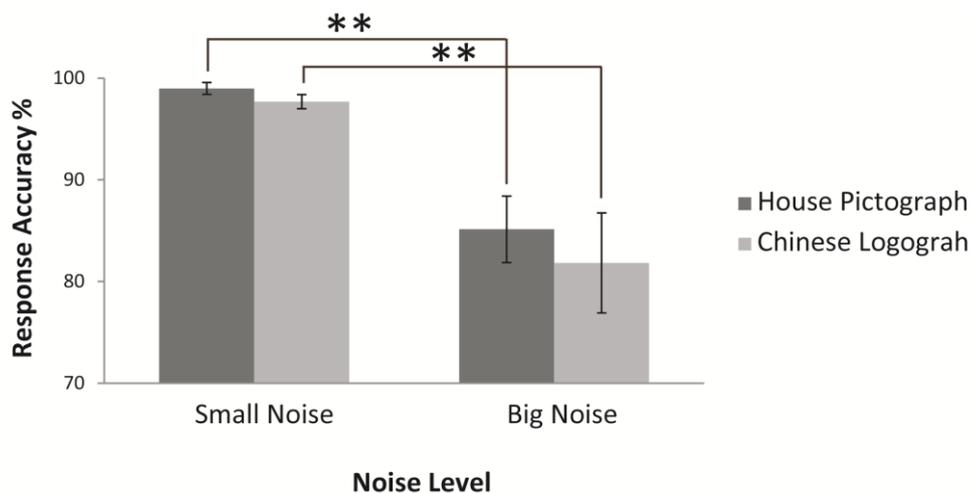


Figure 3.2. Response accuracy for the Chinese characters and house pictographs of the subjects during the fMRI experiment. Two-way repeated-measures ANOVA demonstrated a significant main effect of the noise level ($P < 0.0001$). Pairwise comparisons with Bonferroni corrections showed significant differences between the noise levels within each type of stimulus (* $p < 0.05$, *** $p < 0.0001$, error bars, \pm SEM).

In the VWFA (Figure 3.4.A), there was a significant main effect of noise level (big level of

noise > small level of noise, $F(1, 10)=9.019$, $p<0.05$). No significant main effect of the category or category \times noise level interaction was found (Figure 3.4.B). Pairwise comparisons with Bonferroni corrections showed that a significant activation difference existed between the Chinese characters and the house pictographs with small level of noise [$F(1, 10)=11.217$, $p<0.007$] and between the house pictographs with big and small level of noise [$F(1, 10)=14.022$, $p<0.004$].

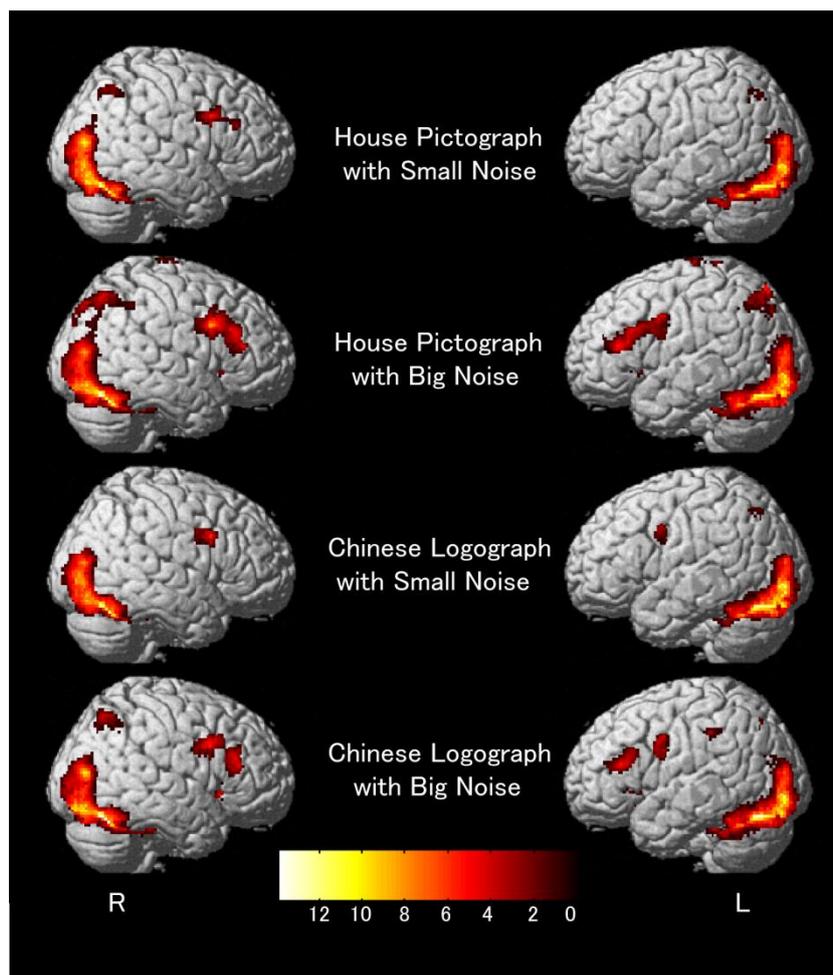


Figure 3.3. Group result (13 subjects) of the whole activation for the four experimental

conditions (Chinese character with small noise, house pictograph with small noise, Chinese character with big noise, house pictograph with big noise) relative to fixation. The activation intensity is indicated by regions in red to yellow. A voxel threshold of $P < 0.05$, corrected with FDR, and a minimum cluster size of 100 mm³ were used.

Percent Signal Change in the VWFA

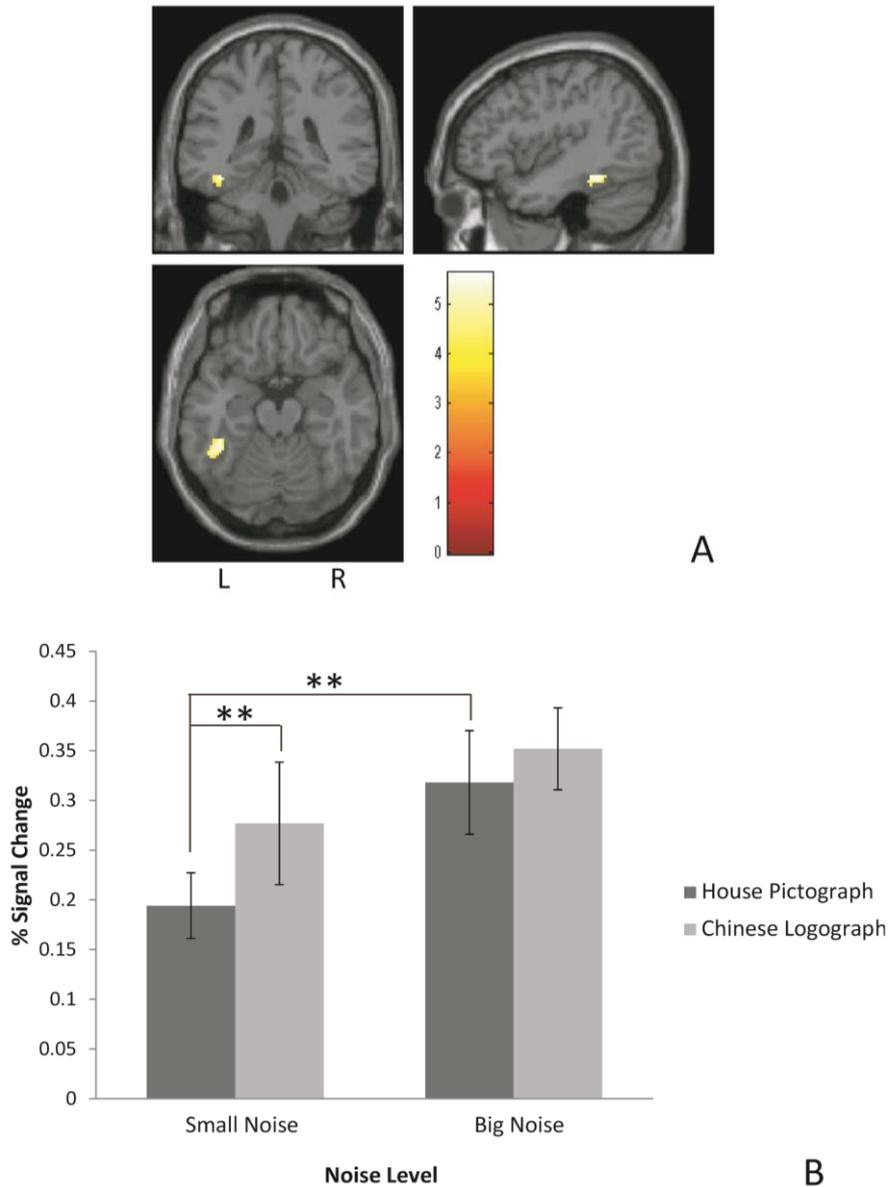


Figure 3.4. The BOLD percentage signal change in the VWFA. (A) Section view of the group VWFA identified in localizer scan. (B) Two-way repeated-measures ANOVA demonstrated a significant main effect of the noise level in the VWFA ($p < 0.05$). Pairwise comparisons with Bonferroni corrections showed significant differences between the Chinese characters with

small noise and the house pictographs with small noise ($p < 0.001$) and also between the house pictographs with big and small noise ($p < 0.001$). (* $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$, error bars, \pm SEM).

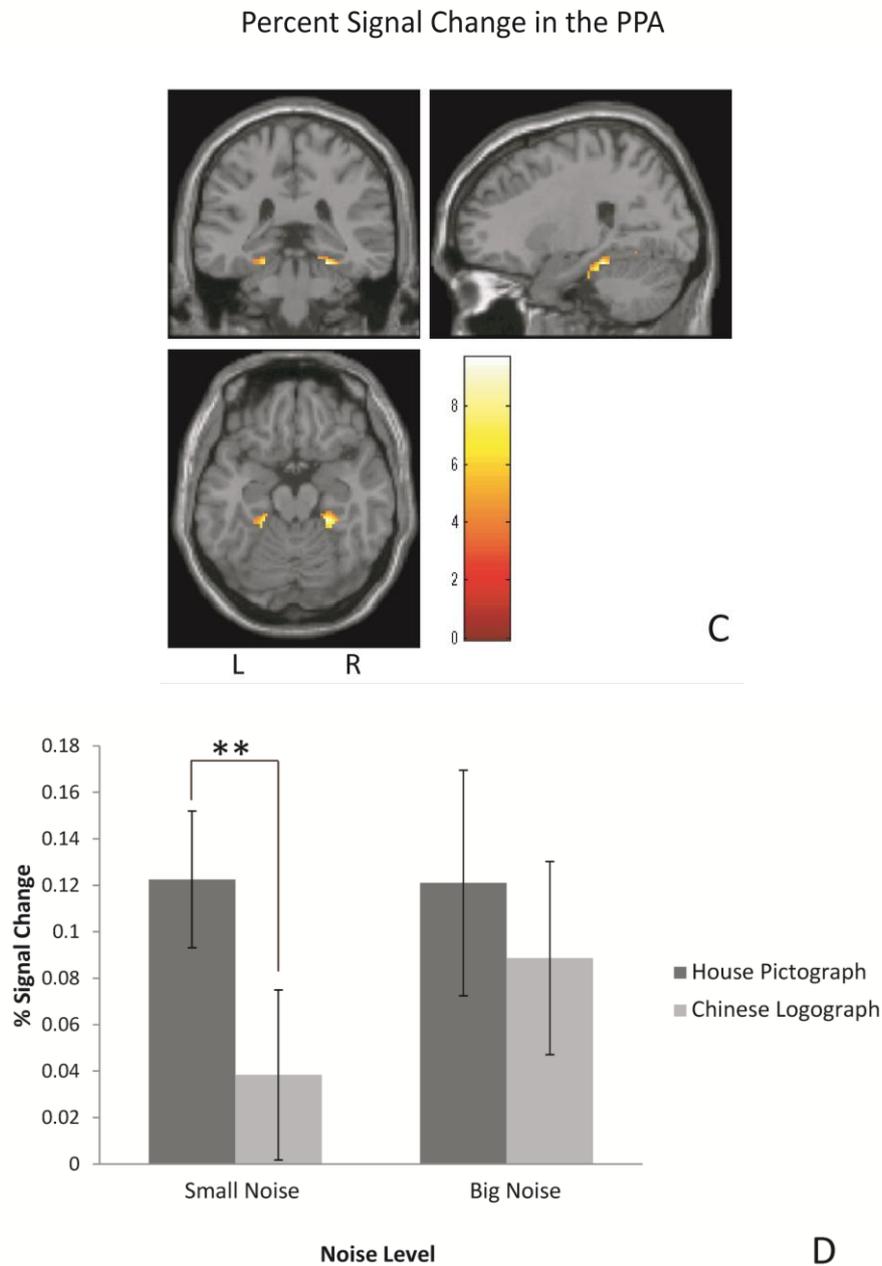


Figure 3.4. The BOLD percentage signal change in the PPA. (C) Section view of the group PPA identified in localizer scan. (D) Two-way repeated-measures ANOVA demonstrated a

significant main effect of category in the PPA ($p < 0.0001$). Pairwise comparisons with Bonferroni corrections showed significant differences between the house pictographs with small noise and the Chinese characters with small noise ($p < 0.001$) (* $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$, error bars, \pm SEM).

In the PPA (Figure 3.4.C), there was a significant effect of category (house pictograph > Chinese character, $F(1,12)=28.416$, $p < 0.0001$). No significant main effect of noise level or category \times noise level interaction was found (Figure 3.4.D). Pairwise comparisons with Bonferroni corrections showed that a significant activation difference existed between the house pictograph and the Chinese character with small level of noise [$F(1, 12)=14.710$, $p < 0.002$].

3.4 Discussion

The aim of this study was to investigate and compare task/attentional modulation in the VWFA and PPA. The influence of small and big noise levels on BOLD response in the VWFA and PPA was examined during Chinese character and house pictograph discrimination task. Previous fMRI studies of object selectivity have typically used naturalistic images of objects that vary in surface properties such as texture, luminance and material. However, these visual features are atypical of visual word stimuli. Comparing naturalistic images with words, therefore, makes it difficult to elucidate the effects of shape from those due to variations in surface features. To avoid this problem, house pictographs and Chinese characters were used as visual stimuli, both of which had similar big spatial frequency square configurations and a uniform gray color. The stimuli with a small level of noise revealed 95% of the sensory

evidence, and the stimuli with a big level of noise revealed 75% of the sensory evidence, making them either easier or more difficult to identify as a house or a word. Repeated measures two-way ANOVA revealed a significant main effect of category but not noise level in the PPA response. This result is consistent with the findings from a previous study by Horner and colleagues (Horner & Andrews, 2009), which indicated that the house-selective response in the PPA was apparent even under the small visibility condition that revealed 75% sensory evidence. The VWFA response, in contrast, demonstrated a significant main effect of the noise level but not category. This finding suggests that the VWFA response was modified by the noise embedded in the stimuli, which demanded more visual attention by increasing the task difficulty.

As can be seen in Figure 3.3, big noise elicited larger and stronger activity compared to small noise in regard to both Chinese logographs and house pictographs. The dorsolateral frontal region has been implicated in executive functions and suggested to be especially interconnected with brain regions involved with attention (Goldmanrakis, 1988; Alvarez & Emory, 2006). This is consistent with the fact that the big noise condition in the present study required more visual attention to correctly execute the task. In addition, there was a predominance of bilateral dorsolateral frontal activations when the noise was bigger. It may reflect the recruitment of homologous resources across hemispheres due to high demands

imposed by task complexity (Kaller et al., 2011).

Visually presented words or letter strings preferentially activate a small region in the left fusiform gyrus (VWFA) (Price, 2000; Jobard et al., 2003; Dehaene et al., 2005). Houses and buildings, on the other hand, typically activate a subregion of the PPA that plays an important role in the encoding and recognition of scenes and houses (rather than faces or objects) (Aguirre et al., 1998; Ishai et al., 1999; R. Epstein et al., 1999). Consistent with these previous findings, the VWFA in the present data showed significantly stronger activation to Chinese characters with small level of noise compared to house pictographs with small level of noise (Figure 3.4B), and the PPA showed preferential activation to house pictographs with small level of noise compared to Chinese characters with small level of noise (Figure 3.4D).

The stronger VWFA sensitivity to Chinese characters may be a consequence of the level of experience with the stimulus category but it may also stem from stimulus features that distinguish Chinese characters from house pictograph. Several lines of cognitive evidence indicated that the VWFA sensitivity is significantly higher to words than to other similarly complex stimulus types, such as line drawings and false-font strings (Ben-Shachar et al., 2007; Baker et al., 2007). Previous studies demonstrated the VWFA sensitivity to the orthographic regularities by which letters form words (Beauregard et al., 1997; Fiebach et al., 2002; Cohen et al., 2002; Dehaene et al., 2004). They found that novel, but well-structured letter strings (i.e.

pseudowords), produce more robust fMRI responses than novel but poorly structured letter strings. Based on these evidences, McCandliss and colleagues (McCandliss et al., 2003) proposed that words are first processed in ventral occipital regions (V1 to V4) contralateral to the stimuli, building up increasingly abstract visual representations and eventually converging in a structure within the VWFA. Subsequent findings provided further support for the visual word form extraction in this region (Binder et al., 2006; Vinckier et al., 2007). More recent studies indicated that the word sensitivity in the VWFA can be explained by an interactive view (Price & Devlin, 2011; Price, 2012). They suggested that the VWFA integrates visuospatial features abstracted from sensory inputs with higher level associations such as speech sounds, actions and meanings. Thus, the stronger VWFA sensitivity to Chinese characters under small noise condition may be explained by the region's critical role in linking character shapes and their associated phonological content (Hashimoto & Sakai, 2004; Striem-Amit et al., 2012).

Unlike the VWFA, the PPA showed stronger response to house pictographs under both noise level conditions. The data can be interpreted as supporting a role of PPA in geometric structure encoding. Both the Chinese characters and house pictographs used in this study had similar square configurations. However, they differed in their spatial distributions and could be divided into two groups of distinct geometric shapes, one of which represented a house and the other a word. Although the spatial details of each stimulus were blurred with the embedded

white noise, the overall geometric structures of the stimuli were still detectable (95% and 75% sensory evidence). Thus, the main effect of category in the PPA response suggests that the region is tuned for the processing of integrated shapes of visual stimuli rather than for detail-oriented processing. This suggestion is consistent with the findings of previous fMRI studies (R. Epstein et al., 2001; Bar & Aminoff, 2003; R. A. Epstein, 2008; Henderson et al., 2008), which have revealed that the PPA encodes the geometric structure of the entire space, integrating the component parts into a global geometric shape that represents the scene, landscape and house.

In contrast to the stimuli with small level of noise, neither the VWFA nor the PPA showed significant category-selective activation for the stimuli with big level of noise. This finding suggests that the stimuli with a big level of noise contain less sensory evidence compared to the stimuli with small level of noise, therefore reducing the difference between the responses to the two categories of stimuli. This is consistent with previous studies (Heekeren et al., 2004; Heekeren et al., 2008; Horner & Andrews, 2009) whose results indicated a decrease in the category-selective response in the VWFA along with an increase in the noise level embedded in the visual stimuli.

Although the stronger VWFA response for the Chinese characters under small level noise condition can be explained by word-selectivity to some extent, the overall response showed a

main effect of the noise level but not the category (Figure 3.4.A). It suggests that the VWFA was equally involved in processing the two types of stimuli. Both categories of stimuli in this experiment were characterized by edges, borders, and small spatial details. Thus, to recognize the stimuli depends not only on processing their gross geometric shape but also on processing their fine-grained visual information.

Malach and colleagues (Levy et al., 2001; Hasson et al., 2002; Malach et al., 2002) have proposed the notion of a retinotopically organized extrastriate cortex. According to this view, there is a graded division of labor across the extrastriate cortex such that fine-grained demanding visual categories (i.e., words) dominate activation of the VWFA. Visual acuity is highest in the fovea and drops toward the parafoveal region (Fiset et al., 2006; Starrfelt et al., 2009). In line with this notion, a recent study on patients with fusiform gyrus damage (Roberts et al., 2012) revealed that the functional role of the VWFA is not limited to fine-grained visual processing of orthographic stimuli but also addresses visually complex stimuli in general. According to the data, the VWFA did not indicate category-selective response to the stimuli under big noise level condition. It suggests that the VWFA was involved in fine visual discrimination of both Chinese characters and house pictographs when the stimuli were more difficult to identify.

The data also showed that the stimuli with big level of noise elicited a stronger response in

the VWFA than the stimuli with small level of noise. This finding suggests that the big noise level in the stimuli increased the task demand, which in turn increased the neural activity in this region for fine visual discrimination. Attention can also account for the increased response to the stimuli with big level of noise. Previous fMRI studies (Wojciulik et al., 1998; Murray & Wojciulik, 2004) have indicated attentional modulation of the magnitude of BOLD responses, showing enhanced activations for attended compared with less-attended visual stimuli. Thus, when the task requires a big level of attention directed to the stimuli, the VWFA is activated more to provide detailed visual information for fine judgments about the stimuli.

3.5 Conclusions

The results showed that the BOLD response in the VWFA and PPA are modulated differently by visual attention. More specifically, the selective response in the VWFA is more easily affected by attentional demand than that in the PPA. These results may suggest that visual attention is distributed differently along the extrastriate cortex.

Chapter 4

General Discussion and General Conclusion

Summary

In this chapter, findings from the previously described two studies are discussed and a general conclusion is made.

4.1 General Discussion

Reading is such a casual and automatic act in our daily life that we tend to forget what a remarkable skill it is and what convenience it has brought along to us. Whether you are engaged in a novel, pouring over a newspaper or just looking at a sign, reading skills allow you to interpret and become engaged in the modern world around you. It is universally known that reading usually refers to converting the visual information of words into their corresponding sounds and meanings, and it can take years of training before we become fluent at it. We may have all at one point in our lives wondered the same question: how do we read? Advances in psychology and neuroscience over the last twenty years have shed light on the underlying mechanism of this amazing feat. Modern brain imaging technologies in particular have enabled us to visualize brain activation during reading in vivo, thereby revealing a chain of word processing stages.

In 2000, a group of French researchers led by Laurent Cohen and Stanislas Dehaene reported a case of word processing in the brain of patients suffering from left hemialexia following posterior callosal lesions. Their study highlighted the damaged region locating at the mid portion of left fusiform gyrus and hypothesized that this area is functionally specialized for processing of “visual word form” – a neural representation after primary visual processing and before phonological and semantic processing. This finding echoed the 1892 discovery of French

neurologist Joseph-Jules Déjerine, who reported a selective impairment of word recognition in patient Mr. C after a stroke in the left visual brain system. Symptoms of the patients described above are very similar to those of people with pure alexia. More specifically, these patients have simply lost their ability to read while they could still normally speak, recognize faces and objects and even write. Normal people, on the other hand, have demonstrated consistent response properties to presented word stimuli in this region. For example, the region responds invariantly to words across a range of retinal positions, sizes and fonts. Laurent Cohen and colleagues, therefore, tentatively labeled this region of interest as the “visual word form area, VWFA”.

Since then, the function of this left fusiform gyrus region (also referred to as left ventral occipito-temporal region, this term will be used in the rest of the content) became the center of a debate marathon for the last decade. The debate has revolved around one major issue – functional specialization. Written words have been in use for only a few thousand years, which is not long enough for natural selection to produce a genetically specified brain region dedicated to visual word recognition. Under such circumstance, why is this region functionally tuned to word processing? Is it involved in recognition of other objects thereby subserving a general perceptual processing? Does this region functionally differ between people who can read and those who can't? To find answers to these questions, a series of studies were performed using

functional magnetic resonance imaging (fMRI) techniques. These studies will be introduced in a chronological order in the following paragraphs.

In the first study, brain activations of literate and illiterate people were compared while they were judging whether two Chinese characters presented simultaneously on the screen were the same or not. 13 subjects were recruited for each study group. In every experimental trial, two random Chinese characters would be shown on the left and right side of a fixation point. The subjects were instructed to press one button if they happen to be the same and press the other button if otherwise. Such cognitive task is usually referred to as perceptual matching. There were also trials for simple figure matching which were utilized as control condition. A significant factor that differentiates the two subject groups was the ability to read. Since the literate group went through years of schooling, they were experts in recognizing Chinese characters. One might expect that such visual word expertise in literate people would facilitate their performance on Chinese character matching task. Indeed, behavioral results showed that although performances in simple figure matching were equally well between the two groups, the literate group outscored the illiterate group during character matching with significantly higher accuracy and faster speed. Then their task related brain activity was further examined to look for differences between the two groups, especially in the ventral occipito-temporal (vOT) region whose functional specialization for words has been debated. Consistent with the expectation, the

illiterate group showed more extensive activation in the right hemisphere whereas the literate group showed more extensive activation in the left hemisphere, suggesting different strategies adopted by each subject group to complete the task. When the occipito-temporal regions was examined closely, however, no significant response difference was found between the two. In a 2010 study conducted by Dehaene and colleagues also compared brain activation between literate and illiterate people by utilizing a passive viewing task. They found that vOT response to written words was significantly stronger in literate compared to the illiterate subjects. The seemingly contradictory results between these two studies may be explained by findings from recent neuroimaging researches, which suggested that vOT response properties are highly dependent upon the task demand. Unlike perceptual matching, a passive viewing task triggers the highly automatic word processing bias in literate subjects, thereby eliciting stronger activation in the vOT. Together these findings suggest that a direct relationship between visual word expertise and enhanced vOT response to written words should be taken with caution, since it can be affected by other factors such as task demand and attention.

In the second study, the effect of task demand and visual attention on vOT response to words was further examined. Here we recruited 13 healthy college students to participate in an fMRI experiment. During the experiment, subjects were asked to decide whether an image presented randomly on screen was a Chinese character or house pictograph. Note that all images

were imbedded with white noise, and half of them with a small level of noise while the other half with a big level of noise. These two levels were assessed in a psychophysical calibration prior to the fMRI experiment. The small level of noise allowed the subjects to correctly discriminate the image with a chance of 95%, whereas the big level of noise only allowed 75% chance of recognition, suggesting an increased difficulty to complete the task. First, brain activations related to Chinese character and house pictograph processing were examined respectively, and localized two distinct areas located along the extrastriate cortex. The area related to Chinese character processing corresponded to the hotly debated VWFA/vOT, and the other area related to house pictograph processing corresponded to the parahippocampal place area (PPA), a subregion of the parahippocampal cortex that plays an important role in the encoding and recognition of scenes. Then, how the two different levels of noise affected activations in these two areas was investigated. Interestingly, they demonstrated different sensitivity to the task demand with the VWFA/vOT more prone to change in the noise level than the PPA. When the whole brain activation was examined, it was found that regardless of character and house, the big noise level elicited more extensive and stronger response in areas involved in attention and executive function compared to the small noise level. These findings suggest that big noise level increased the task difficulty, and therefore more attention was urged to make the right judgment. Besides word- and house/scene-selective response properties,

previous studies also demonstrated that the VWFA/vOT is more sensitive to fine-grained visual information, whereas the PPA tends to be more sensitive to the gross geometrical shape of object. When an image of object is blurred by white noise, it usually becomes more difficult to discriminate the visual details of the object than its contours. To make the right judgment, more fine-grained information is needed to accurately identify the object. Thus the heavy responsibility falls down to the VWFA/vOT, making its response subject to change. In the case of the second study, the cause of this change was increased task difficulty due to increased noise in the stimuli. Therefore, different sensitivity to task demand in the VWFA/vOT and the PPA could be viewed as dispersed distribution of attention based on their different functional properties. This study provided us another evidence that the response property of the VWFA/vOT is apt to change with task demand and level of attention, therefore underlining the importance to look more at the region's functional plasticity.

4.2 General Conclusion

To summarize the findings from the two studies: firstly, the vOT responses to visual words are similar between literate and illiterate adults during perceptual matching task, suggesting this region may be involved in general visuospatial processing; secondly, the vOT is easily affected by increase in the noise level, suggesting the region may be involved in visual processing of both word and house stimuli. Overall, the vOT is not specialized in visual word processing, but

is involved in other visual processing as well.

In the future study, which is already an ongoing research, it is focus to explore functional plasticity of the vOT by investigating its multimodal response properties. Although previous imaging researches have predominantly studies the vOT involvement in visual word recognition, more and more evidences have started to show that word processing in the vOT is not modality specific. The strongest evidences came from studying blind subjects reading Braille words, which indicated that people without visual experiences were able to recruit the vOT for tactile word recognition so that it functioned equally as the VWFA in sighted subjects. Another profound evidence for functional plasticity of this region was provided by a study in which subjects were trained to “hear” the shapes of letters in with soundscape, and subsequently activating the vOT for the discrimination. In the light of these and other related findings, the hypothesis is that the vOT converge information from different modality input and communicate them with higher-level language areas. To test this hypothesis, a behavioral experiment have been conducted to examine cross modality priming between visual, auditory and tactile word processing. The results showed that the priming effect existed between both within and across modality. The next step is to conduct an fMRI experiment to see cross modality priming effect in the vOT.

Findings from these studies and those from other research group have expanded the

outlook of this particular region – vOT by revealing its variety of functional properties which are beyond just visual word processing. These discoveries helped answering some of the initial questions about the region while others remain to be explored. We are still far from a perfect understanding, and cognitive models based on word/object recognition are still under construction. However, even with our current level of knowledge, engineers have already started to implement the same algorithms in mechanical devices to improve their performance. In the sphere of clinical application, understanding of brain mechanism for word recognition has also helped to develop rehabilitation method for children with dyslexia. Beyond these important applications, understanding why we are able to read and write may also become a key to see what makes us human, since we are the only species capable of developing such ability. The next thing we should do to proceed on this path is to gain insights from workings of other brain areas as well, especially the object-selective regions (i.e. PPA) which locate adjacently along the ventral visual pathway. These regions have been implicated in visual expertise of various object recognitions – the magnificent ability of our visual system to efficiently perceive the world around us. Also how the vOT communicate with other parts of the brain should also be a focus of study by combining both spatial and temporal account of the processing flow.

Appendices

I. Introduction to fMRI

Functional magnetic resonance imaging or functional MRI (fMRI) is a functional neuroimaging procedure using MRI technology (Figure I) that measures brain activity by detecting associated changes in blood flow. This technique relies on the fact that cerebral blood flow and neuronal activation are coupled. When an area of the brain is in use, blood flow to that region also increases.



Figure I. Front and side view of a Philips Ingenia 3T scanner. MRI scanners use strong magnetic fields and radiowaves to form images of the body. The technique is widely used in hospitals for medical diagnosis, staging of disease and for follow-up without exposure to ionizing radiation.

The primary form of fMRI uses the Blood-oxygen-level dependent (BOLD) contrast, discovered by Seiji Ogawa. This is a type of specialized brain and body scan used to map neural activity in the brain or spinal cord of humans or other animals by imaging the change in blood

flow (hemodynamic response) related to energy use by brain cells. Since the early 1990s, fMRI has come to dominate brain mapping research because it does not require people to undergo shots, surgery, or to ingest substances, or be exposed to radiation, etc. Other methods of obtaining contrast are arterial spin labeling and diffusion MRI.

The procedure is similar to MRI but uses the change in magnetization between oxygen-rich and oxygen-poor blood as its basic measure. This measure is frequently corrupted by noise from various sources and hence statistical procedures are used to extract the underlying signal. The resulting brain activation can be presented graphically by color-coding the strength of activation across the brain or the specific region studied. The technique can localize activity to within millimeters but, using standard techniques, no better than within a window of a few seconds.

fMRI is used both in the research world, and to a lesser extent, in the clinical world. It can also be combined and complemented with other measures of brain physiology such as EEG and NIRS. Newer methods which improve both spatial and time resolution are being researched, and these largely use biomarkers other than the BOLD signal. Some companies have developed commercial products such as lie detectors based on fMRI techniques, but the research is not believed to be ripe enough for widespread commercialization.

The fMRI concept builds on the earlier MRI scanning technology and the discovery of properties of oxygen-rich blood. MRI brain scans use a strong, permanent, static magnetic field

to align nuclei in the brain region being studied. Another magnetic field, the gradient field, is then applied to kick the nuclei to higher magnetization levels, with the effect depending on where they are located. When the gradient field is removed, the nuclei go back to their original states, and the energy they emit is measured with a coil to recreate the positions of the nuclei. MRI thus provides a static structural view of brain matter. The central thrust behind fMRI was to extend MRI to capture functional changes in the brain caused by neuronal activity. Differences in magnetic properties between arterial (oxygen-rich) and venous (oxygen-poor) blood provided this link.

Since the 1890s it has been known that changes in blood flow and blood oxygenation in the brain (collectively known as hemodynamics) are closely linked to neural activity. When neurons become active, local blood flow to those brain regions increases, and oxygen-rich (oxygenated) blood displaces oxygen-depleted (deoxygenated) blood around 2 seconds later. This rises to a peak over 4–6 seconds, before falling back to the original level (and typically undershooting slightly). Oxygen is carried by the hemoglobin molecule in red blood cells. Deoxygenated hemoglobin (dHb) is more magnetic (paramagnetic) than oxygenated hemoglobin (Hb), which is virtually resistant to magnetism (diamagnetic). This difference leads to an improved MR signal since the diamagnetic blood interferes with the magnetic MR signal less. This improvement can be mapped to show which neurons are active at a time.

II. BOLD Hemodynamic Response

The change in the MR signal from neuronal activity is called the hemodynamic response (HDR). It lags the neuronal events triggering it by 1 to 2 seconds, since it takes that long for the vascular system to respond to the brain's need for glucose. From this point it typically rises to a peak at about 5 seconds after the stimulus. If the neurons keep firing, say from a continuous stimulus, the peak spreads to a flat plateau while the neurons stay active. After activity stops, the BOLD signal falls below the original level, the baseline, a phenomenon called the undershoot. Over time the signal recovers to the baseline. There is some evidence continuous metabolic requirements in a brain region contribute to the undershoot.

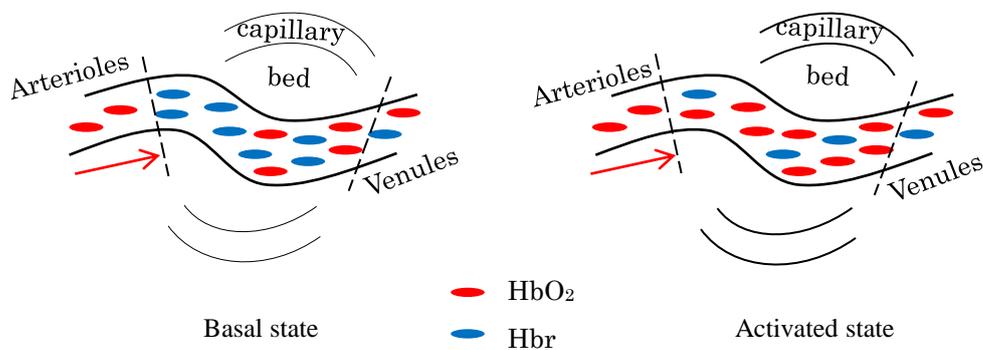


Figure II. Neurons do not have internal reserves of energy in the form of sugar and oxygen, so their firing causes a need for more energy to be brought in quickly. Through a process called the hemodynamic response, blood releases oxygen to them at a greater rate than to inactive neurons, and the difference in magnetic susceptibility between oxyhemoglobin and deoxyhemoglobin, and thus oxygenated or deoxygenated blood, leads to magnetic signal variation which can be detected using an MRI scanner. HbO₂: oxyhemoglobin, Hbr: deoxyhemoglobin.

The mechanism by which the neural system provides feedback to the vascular system of its need for more glucose is partly the release of glutamate as part of neuron firing. This glutamate affects nearby supporting cells, astrocytes, causing a change in calcium ion concentration. This, in turn, releases nitric oxide at the contact point of astrocytes and intermediate-sized blood vessels, the arterioles. Nitric oxide is a vasodilator causing arterioles to expand and draw in more blood.

A single voxel's response signal over time is called its timecourse. Typically, the unwanted signal called the noise, from the scanner, random brain activity and similar elements, is as big as the signal itself. To eliminate these, fMRI studies repeat a stimulus presentation multiple times.

Spatial resolution of an fMRI study refers to how well it discriminates between nearby locations. It is measured by the size of voxels, as in MRI. A voxel is a three-dimensional rectangular cuboid, whose dimensions are set by the slice thickness, the area of a slice, and the grid imposed on the slice by the scanning process. Full-brain studies use larger voxels, while those that focus on specific regions of interest typically use smaller sizes. Sizes range from 4 to 5 mm to 1 mm. Smaller voxels contain fewer neurons on average, incorporate less blood flow, and hence have less signal than larger voxels. Smaller voxels also take longer to scan, since scanning time directly rises with the number of voxels per slice and the number of slices. This can lead both to discomfort for the subject inside the scanner and to loss of the magnetization

signal. A voxel typically contains a few million neurons and tens of billions of synapses, with the actual number depending on voxel size and the area of the brain being imaged.

The vascular arterial system supplying fresh blood branches into smaller and smaller vessels as it enters the brain surface and within-brain regions, culminating in a connected capillary bed within the brain. The drainage system, similarly, merges into larger and larger veins as it carries away oxygen-depleted blood. The dHb contribution to the fMRI signal is from both the capillaries near the area of activity and larger draining veins that may be farther away. For good spatial resolution, the signal from the large veins needs to be suppressed, since it does not correspond to the area where the neural activity is. This can be achieved either by using strong static magnetic fields or by using spin-echo pulse sequences. With these, fMRI can examine a spatial range from millimeters to centimeters, and can hence identify Brodmann areas (centimeters), subcortical nuclei such as the caudate, putamen and thalamus, and hippocampal subfields such as the combined dentate gyrus/CA3, CA1, and subiculum.

Temporal resolution is the smallest time period of neural activity reliably separated out by fMRI. One element deciding this is the sampling time, the TR. Below a TR of 1 or 2 seconds, however, scanning just generates sharper HDR curves, without adding much information beyond what mathematically filling in the gaps of the curve produced by a higher TR would. Temporal resolution can be improved by staggering stimulus presentation across trials. If

one-third of data trials are sampled normally, one-third at 1 s, 4 s, 7 s and so on, and the last third at 2 s, 5 s and 8 s, the combined data provide a resolution of 1 s, though with only one-third as many total events.

The time resolution needed depends on brain processing time for various events. An example of the broad range here is given by the visual processing system. What the eye sees is registered on the photoreceptors of the retina within a millisecond or so. These signals get to the primary visual cortex via the thalamus in tens of milliseconds. Neuronal activity related to the act of seeing lasts for more than 100 ms. A fast reaction, such as swerving to avoid a car crash, takes around 200 ms. By about half-a-second, awareness and reflection of the incident sets in. Remembering a similar event may take a few seconds, and emotional or physiological changes such as fear arousal may last minutes or hours. Learned changes, such as recognizing faces or scenes, may last days, months, or forever. Most fMRI experiments study brain processes lasting a few seconds, with the study conducted over some tens of minutes. Subjects may move their heads during that time, and this head motion needs to be corrected for. So does drift in the baseline signal over time. Boredom and learning may modify both subject behavior and cognitive processes.

When a person performs two tasks simultaneously or in overlapping fashion, the BOLD response is expected to add linearly. This is a fundamental assumption of many fMRI studies.

Linear addition means the only operation allowed on the individual responses before they are combined (added together) is a separate scaling of each. Since scaling is just multiplication by a constant number, this means an event that evokes, say, twice the neural response as another, can be modeled as the first event presented twice simultaneously. The HDR for the doubled-event is then just double that of the single event.

This strong assumption was first studied in 1996 by Boynton and colleagues, who checked the effects on the primary visual cortex of patterns flickering 8 times a second and presented for 3 to 24 seconds. Their result showed that when visual contrast of the image was increased, the HDR shape stayed the same but its amplitude increased proportionally. With some exceptions, responses to longer stimuli could also be inferred by adding together the responses for multiple shorter stimuli summing to the same longer duration. In 1997, Dale and Buckner tested whether individual events, rather than blocks of some duration, also summed the same way, and found they did. But they also found deviations from the linear model at time intervals less than 2 seconds.

A source of nonlinearity in the fMRI response is from the refractory period, where brain activity from a presented stimulus suppresses further activity on a subsequent, similar, stimulus. As stimuli become shorter, the refractory period becomes more noticeable. The refractory period does not change with age, nor do the amplitudes of HDRs. The period differs across brain

regions. In both the primary motor cortex and the visual cortex, the HDR amplitude scales linearly with duration of a stimulus or response. In the corresponding secondary regions, the supplementary motor cortex, which is involved in planning motor behavior, and the motion-sensitive V5 region, a strong refractory period is seen and the HDR amplitude stays steady across a range of stimulus or response durations. The refractory effect can be used in a way similar to habituation to see what features of a stimulus a person discriminates as new.

III. Data Analysis with SPM

The goal of fMRI data analysis is to detect correlations between brain activation and a task the subject performs during the scan. It also aims to discover correlations with the specific cognitive states, such as memory and recognition, induced in the subject. The BOLD signature of activation is relatively weak, however, so other sources of noise in the acquired data must be carefully controlled. This means that a series of processing steps must be performed on the acquired images before the actual statistical search for task-related activation can begin. Nevertheless, it is possible to predict, for example, the emotions a person is experiencing solely from their fMRI, with a high degree of accuracy.

Noise is unwanted changes to the MR signal from elements not of interest to the study. The five main sources of noise in fMRI are thermal noise, system noise, physiological noise, random neural activity and differences in both mental strategies and behavior across people and across

tasks within a person. Thermal noise multiplies in line with the static field strength, but physiological noise multiplies as the square of the field strength. Since the signal also multiplies as the square of the field strength, and since physiological noise is a large proportion of total noise, higher field strengths above 3 T do not always produce proportionately better images.

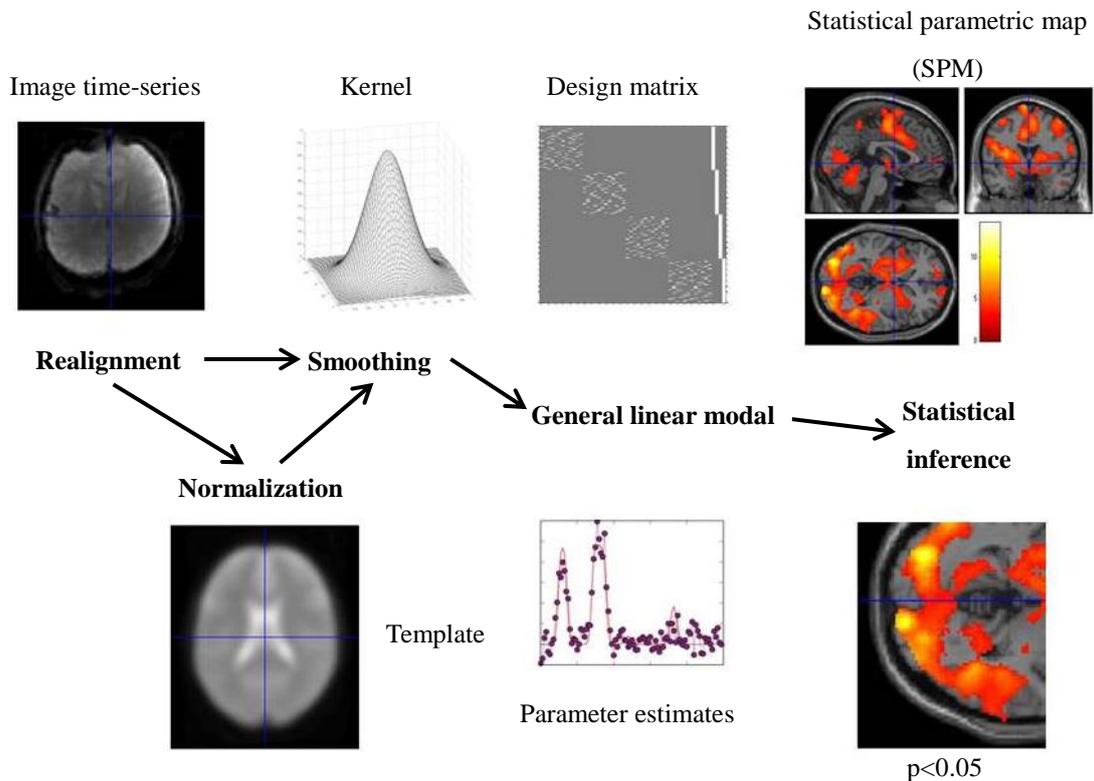


Figure III. This schematic depicts the transformations that start with an imaging data sequence and end with a statistical parametric map (SPM). An SPM can be regarded as an 'X-ray' of the significance of regional effects. Voxel-based analyses require the data to be in the same anatomical space: this is effected by realigning the data. After realignment, the images are subject to non-linear warping so that they match a spatial model or template that already conforms to a standard anatomical space. After smoothing, the general linear model is employed to estimate the parameters of a temporal model (encoded by a design matrix) and derive the appropriate univariate test statistic at every voxel. The test statistics (usually t or F-statistics) constitute the SPM. The final stage is to make statistical inferences on the basis of the SPM and Random Field Theory and characterize the responses observed using the fitted responses or parameter estimates.

Heat causes electrons to move around and distort the current in the fMRI detector, producing thermal noise. Thermal noise rises with the temperature. It also depends on the range of frequencies detected by the receiver coil and its electrical resistance. It affects all voxels similarly, independent of anatomy.

System noise is from the imaging hardware. One form is scanner drift, caused by the superconducting magnet's field drifting over time. Another form is changes in the current or voltage distribution of the brain itself inducing changes in the receiver coil and reducing its sensitivity. A procedure called impedance matching is used to bypass this inductance effect. There could also be noise from the magnetic field not being uniform. This is often adjusted for by using shimming coils, small magnets physically inserted, say into the subject's mouth, to patch the magnetic field. The nonuniformities are often near brain sinuses such as the ear and plugging the cavity for long periods can be discomfiting. The scanning process acquires the MR signal in k-space, in which overlapping spatial frequencies (that is repeated edges in the sample's volume) are each represented with lines. Transforming this into voxels introduces some loss and distortions.

Physiological noise is from head and brain movement in the scanner from breathing, heart beats, or the subject fidgeting, tensing, or making physical responses such as button presses. Head movements cause the voxel-to-neurons mapping to change while scanning is in progress.

Since fMRI is acquired in slices, after movement, a voxel continues to refer to the same absolute location in space while the neurons underneath it would have changed. Another source of physiological noise is the change in the rate of blood flow, blood volume, and use of oxygen over time. This last component contributes to two-thirds of physiological noise, which, in turn, is the main contributor to total noise.

Even with the best experimental design, it is not possible to control and constrain all other background stimuli impinging on a subject—scanner noise, random thoughts, physical sensations, and the like. These produce neural activity independent of the experimental manipulation. These are not amenable to mathematical modeling and have to be controlled by the study design.

A person's strategies to respond or react to a stimulus, and to solve problems, often change over time and over tasks. This generates variations in neural activity from trial to trial within a subject. Across people too neural activity differs for similar reasons. Researchers often conduct pilot studies to see how participants typically perform for the task under consideration. They also often train subjects how to respond or react in a trial training session prior to the scanning one.

Preprocessing

The scanner platform generates a 3 D volume of the subject's head every TR. This consists

of an array of voxel intensity values, one value per voxel in the scan. The voxels are arranged one after the other, unfolding the three-dimensional structure into a single line. Several such volumes from a session are joined together to form a 4 D volume corresponding to a run, for the time period the subject stayed in the scanner without adjusting head position. This 4 D volume is the starting point for analysis. The first part of that analysis is preprocessing.

The first step in preprocessing is conventionally slice timing correction. The MR scanner acquires different slices within a single brain volume at different times, and hence the slices represent brain activity at different timepoints. Since this complicates later analysis, a timing correction is applied to bring all slices to the same timepoint reference. This is done by assuming the timecourse of a voxel is smooth when plotted as a dotted line. Hence the voxel's intensity value at other times not in the sampled frames can be calculated by filling in the dots to create a continuous curve.

Head motion correction is another common preprocessing step. When the head moves, the neurons under a voxel move and hence its timecourse now represents largely that of some other voxel in the past. Hence the timecourse curve is effectively cut and pasted from one voxel to another. Motion correction tries different ways of undoing this to see which undoing of the cut-and-paste produces the smoothest timecourse for all voxels. The undoing is by applying a rigid-body transform to the volume, by shifting and rotating the whole volume data to account

for motion. The transformed volume is compared statistically to the volume at the first timepoint to see how well they match, using a cost function such as correlation or mutual information. The transformation that gives the minimal cost function is chosen as the model for head motion. Since the head can move in a vastly varied number of ways, it is not possible to search for all possible candidates; nor is there right now an algorithm that provides a globally optimal solution independent of the first transformations we try in a chain.

Distortion corrections account for field nonuniformities of the scanner. One method, as described before, is to use shimming coils. Another is to recreate a field map of the main field by acquiring two images with differing echo times. If the field were uniform, the differences between the two images also would be uniform. Note these are not true preprocessing techniques since they are independent of the study itself. Bias field estimation is a real preprocessing technique using mathematical models of the noise from distortion, such as Markov random fields and expectation maximization algorithms, to correct for distortion.

In general, fMRI studies acquire both many functional images with fMRI and a structural image with MRI. The structural image is usually of a higher resolution and depends on a different signal, the T1 magnetic field decay after excitation. To demarcate regions of interest in the functional image, one needs to align it with the structural one. Even when whole-brain analysis is done, to interpret the final results, that is to figure out which regions the active voxels

fall in, one has to align the functional image to the structural one. This is done with a coregistration algorithm that works similar to the motion-correction one, except that here the resolutions are different, and the intensity values cannot be directly compared since the generating signal is different.

Typical MRI studies scan a few different subjects. To integrate the results across subjects, one possibility is to use a common brain atlas, and adjust all the brains to align to the atlas, and then analyze them as a single group. The atlases commonly used are the Talairach one, a single brain of an elderly woman created by Jean Talairach, and the Montreal Neurological Institute (MNI) one. The second is a probabilistic map created by combining scans from over a hundred individuals. This normalization to a standard template is done by mathematically checking which combination of stretching, squeezing, and warping reduces the differences between the target and the reference. While this is conceptually similar to motion correction, the changes required are more complex than just translation and rotation, and hence optimization even more likely to depend on the first transformations in the chain that is checked.

Temporal filtering is the removal of frequencies of no interest from the signal. A voxel's intensity change over time can be represented as the sum of a number of different repeating waves with differing periods and heights. A plot with these periods on the x-axis and the heights on the y-axis is called a power spectrum, and this plot is created with the Fourier transform

technique. Temporal filtering amounts to removing the periodic waves not of interest to us from the power spectrum, and then summing the waves back again, using the inverse Fourier transform to create a new timecourse for the voxel. A high-pass filter removes the lower frequencies, and the lowest frequency that can be identified with this technique is the reciprocal of twice the TR. A low-pass filter removes the higher frequencies, while a band-pass filter removes all frequencies except the particular range of interest.

Smoothing, or spatial filtering, is the idea of averaging the intensities of nearby voxels to produce a smooth spatial map of intensity change across the brain or region of interest. The averaging is often done by convolution with a Gaussian filter, which, at every spatial point, weights neighboring voxels by their distance, with the weights falling exponentially following the bell curve. If the true spatial extent of activation, that is the spread of the cluster of voxels simultaneously active, matches the width of the filter used, this process improves the signal-to-noise ratio. It also makes the total noise for each voxel follow a bell-curve distribution, since adding together a large number of independent, identical distributions of any kind produces the bell curve as the limit case. But if the presumed spatial extent of activation does not match the filter, signal is reduced.

Statistical analysis

These fMRI images are from a study showing parts of the brain lighting up on seeing

houses and other parts on seeing faces. The 'r' values are correlations, with higher positive or negative values indicating a better match.

One common statistical model used for fMRI data analysis is the univariate GLM model, which analyzes each voxel's data separately. The model assumes, at every time point, the HDR is equal to the scaled and summed version of the events active at that point. A researcher creates a design matrix specifying which events are active at any timepoint. One common way is to create a matrix with one column per overlapping event, and one row per time point, and to mark it with a one if a particular event, say a stimulus, is active at that time point. One then assumes a specific shape for the HDR, leaving only its amplitude changeable in active voxels. The design matrix and this shape are used to generate a prediction of the exact HDR response of the voxel at every timepoint, using the mathematical procedure of convolution. This prediction does not include the scaling required for every event before summing them.

The basic model assumes the observed HDR is the predicted HDR scaled by the weights for each event and then added, with noise mixed in. This generates a set of linear equations with more equations than unknowns. A linear equation has an exact solution, under most conditions, when equations and unknowns match. Hence one could choose any subset of the equations, with the number equal to the number of variables, and solve them. But, when these solutions are plugged into the left-out equations, there will be a mismatch between the right and left sides, the

error. The GLM model attempts to find the scaling weights that minimize the sum of the squares of the error. This method is provably optimal if the error were distributed as a bell curve, and if the scaling-and-summing model were accurate. For a more mathematical description of the GLM model, see generalized linear models.

The GLM model does not take into account the contribution of relationships between multiple voxels. Whereas GLM analysis methods assess whether a voxel or region's signal amplitude is higher or lower for one condition than another, newer statistical models such as multi-voxel pattern analysis (MVPA), utilize that unique contributions of multiple voxels within a voxel-population. In a typical implementation, a classifier or more basic algorithm is trained to distinguish trials for different conditions within a subset of the data. The trained model is then tested by predicting the conditions of the remaining (independent) data. This is most typically achieved by training and testing on different scanner sessions or runs. If the classifier is linear, then the training model is a set of weights used to scale the value in each voxel before summing them to generate a single number that determines the condition for each testing set trial. More information on training and testing classifiers is at statistical classification.

References

- Aguirre, G. K., Zarahn, E., D'Esposito, M., 1998. An area within human ventral cortex sensitive to "building" stimuli: Evidence and implications. *Neuron*, 21(2), 373-383.
- Alvarez, J. A., Emory, E., 2006. Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review*, 16(1), 17-42.
- Baker, C. I., Liu, J., Wald, L. L., Kwong, K. K., Benner, T., Kanwisher, N., 2007. Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 104(21), 9087-9092.
- Bar, M., Aminoff, E., 2003. Cortical analysis of visual context. *Neuron*, 38(2), 347-358.
- Beauregard, M., Chertkow, H., Bub, D., Murtha, S., Dixon, R., Evans, A., 1997. The neural substrate for concrete, abstract, and emotional word lexica: A positron emission tomography study. *Journal of Cognitive Neuroscience*, 9(4), 441-461.
- Ben-Shachar, M., Dougherty, R. F., Deutsch, G. K., Wandell, B. A., 2007. Differential sensitivity to words and shapes in ventral occipito-temporal cortex. *Cerebral Cortex*, 17(7), 1604-1611.
- Ben-Shachar, M., Dougherty, R. F., Deutsch, G. K., Wandell, B. A., 2011. The Development of Cortical Sensitivity to Visual Word Forms. *Journal of Cognitive Neuroscience*, 23(9), 2387-2399.
- Binder, J. R., Medler, D. A., Westbury, C. F., Liebenthal, E., Buchanan, L., 2006. Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage*, 33(2), 739-748.
- Bolger, D. J., Perfetti, C. A., Schneider, W., 2005. Cross-cultural effect on the brain revisited: Universal structures plus writing system variation. *Human Brain Mapping*, 25(1), 92-104.
- Buchel, C., Price, C., Friston, K., 1998. A multimodal language region in the ventral visual pathway. *Nature*, 394(6690), 274-277.
- Carreiras, M., Seghier, M. L., Baquero, S., Estevez, A., Lozano, A., Devlin, J. T., Price, C. J., 2009. An anatomical signature for literacy. *Nature*, 461(7266), 983-U245.
- Cohen, L., Dehaene, S., Naccache, L., Lehericy, S., Dehaene-Lambertz, G., Henaff, M. A., Michel, F., 2000. The visual word form area - Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*,

- 123, 291-307.
- Cohen, L., Lehericy, S., Chochon, F., Lemer, C., Rivaud, S., Dehaene, S., 2002. Language-specific tuning of visual cortex functional properties of the Visual Word Form Area. *Brain*, 125, 1054-1069.
- Cohen, L., Martinaud, O., Lemer, C., Lehericy, S., Samson, Y., Obadia, M., Slachevsky, A., Dehaene, S., 2003. Visual word recognition in the left and right hemispheres: Anatomical and functional correlates of peripheral alexias. *Cerebral Cortex*, 13(12), 1313-1333.
- Dehaene, S., Cohen, L., Sigman, M., Vinckier, F., 2005. The neural code for written words: a proposal. *Trends in Cognitive Sciences*, 9(7), 335-341.
- Dehaene, S., Jobert, A., Naccache, L., Ciuciu, P., Poline, J. B., Le Bihan, D., Cohen, L., 2004. Letter binding and invariant recognition of masked words - Behavioral and neuroimaging evidence. *Psychological Science*, 15(5), 307-313.
- Dehaene, S., Le Clec'H, G., Poline, J. B., Le Bihan, D., Cohen, L., 2002. The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport*, 13(3), 321-325.
- Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J. F., Poline, J. B., Riviere, D., 2001. Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, 4(7), 752-758.
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Nunes, G., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., Cohen, L., 2010. How Learning to Read Changes the Cortical Networks for Vision and Language. *Science*, 330(6009), 1359-1364.
- Epstein, R., DeYoe, E. A., Press, D. Z., Rosen, A. C., Kanwisher, N., 2001. Neuropsychological evidence for a topographical learning mechanism in parahippocampal cortex. *Cognitive Neuropsychology*, 18(6), 481-508.
- Epstein, R., Harris, A., Stanley, D., Kanwisher, N., 1999. The parahippocampal place area: Recognition, navigation, or encoding? *Neuron*, 23(1), 115-125.
- Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. *Nature*, 392(6676), 598-601.
- Epstein, R. A., 2008. Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, 12(10), 388-396.
- Fiebach, C. J., Friederici, A. D., Muller, K., von Cramon, D. Y., 2002. fMRI evidence for dual routes to the mental lexicon in visual word recognition. *Journal of Cognitive Neuroscience*, 14(1), 11-23.
- Fiset, D., Gosselin, F., Blais, C., Arguin, M., 2006. Inducing letter-by-letter dyslexia in normal readers. *Journal of Cognitive Neuroscience*, 18(9), 1466-1476.

- Gaillard, R., Naccache, L., Pinel, P., Clemenceau, S., Volle, E., Hasboun, D., Dupont, S., Baulac, M., Dehaene, S., Adam, C., Cohen, L., 2006. Direct intracranial, fMRI, and lesion evidence for the causal role of left inferotemporal cortex in reading. *Neuron*, 50(2), 191-204.
- Garrett, A. S., Flowers, D. L., Absher, J. R., Fahey, F. H., Gage, H. D., Keyes, J. W., Porrino, L. J., Wood, F. B., 2000. Cortical activity related to accuracy of letter recognition. *Neuroimage*, 11(2), 111-123.
- Gauthier, I., 2000. What constrains the organization of the ventral temporal cortex? *Trends in Cognitive Sciences*, 4(1), 1-2.
- Gauthier, I., Skudlarski, P., Gore, J. C., Anderson, A. W., 2000. Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3(2), 191-197.
- Goldman-Rakic, P. S., 1988. Topography of Cognition - Parallel Distributed Networks in Primate Association Cortex. *Annual Review of Neuroscience*, 11, 137-156.
- Grill-Spector, K., Malach, R., 2004. The human visual cortex. *Annu Rev Neurosci*, 27, 649-677.
- Han, J. H., Choi, W., Chang, Y., Jeong, O. R., Nam, K., 2005. Neuroanatomical analysis for onomatopoeia and phonomime words: fMRI study. *Advances in Natural Computation, Pt 1, Proceedings*, 3610, 850-854.
- Harel, A., Gilaie-Dotan, S., Malach, R., Bentin, S., 2010. Top-Down Engagement Modulates the Neural Expressions of Visual Expertise. *Cerebral Cortex*, 20(10), 2304-2318.
- Hashimoto, R., Sakai, K. L., 2004. Learning letters in adulthood: Direct visualization of cortical plasticity for forming a new link between orthography and phonology. *Neuron*, 42(2), 311-322.
- Hasson, U., Levy, I., Behrmann, M., Hendler, T., Malach, R., 2002. Eccentricity bias as an organizing principle for human high-order object areas. *Neuron*, 34(3), 479-490.
- Heekeren, H. R., Marrett, S., Bandettini, P. A., Ungerleider, L. G., 2004. A general mechanism for perceptual decision-making in the human brain. *Nature*, 431(7010), 859-862.
- Heekeren, H. R., Marrett, S., Ungerleider, L. G., 2008. The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience*, 9(6), 467-479.
- Henderson, J. M., Larson, C. L., Zhu, D. C., 2008. Full Scenes produce more activation than Close-up Scenes and Scene-Diagnostic Objects in parahippocampal and retrosplenial cortex: An fMRI study. *Brain and Cognition*, 66(1), 40-49.
- Horner, A. J., Andrews, T. J., 2009. Linearity of the fMRI Response in Category-Selective Regions of Human Visual Cortex. *Human Brain Mapping*, 30(8), 2628-2640.
- Ishai, A., Ungerleider, L. G., Martin, A., Schouten, H. L., Haxby, J. V., 1999. Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 96(16), 9379-9384.

- Jobard, G., Crivello, F., Tzourio-Mazoyer, N., 2003. Evaluation of the dual route theory of reading: a metaanalysis of 35 neuroimaging studies. *Neuroimage*, 20(2), 693-712.
- Joseph, J. E., Cerullo, M. A., Farley, A. B., Steinmetz, N. A., Mier, C. R., 2006. fMRI correlates of cortical specialization and generalization for letter processing. *Neuroimage*, 32(2), 806-820.
- Jung-Beeman, M., 2005. Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences*, 9(11), 512-518.
- Kaller, C. P., Rahm, B., Spreer, J., Weiller, C., Unterrainer, J. M., 2011. Dissociable Contributions of Left and Right Dorsolateral Prefrontal Cortex in Planning. *Cerebral Cortex*, 21(2), 307-317.
- Kanwisher, N., McDermott, J., Chun, M. M., 1997. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17(11), 4302-4311.
- Levy, I., Hasson, U., Avidan, G., Hendler, T., Malach, R., 2001. Center-periphery organization of human object areas. *Nat Neurosci*, 4(5), 533-539.
- Li, G., Cheung, R. T. F., Gao, J. H., Lee, T. M. C., Tan, L. H., Fox, P. T., Jack, C. R., Yang, E. S., 2006. Cognitive processing in Chinese literate and illiterate subjects: An fMRI study. *Human Brain Mapping*, 27(2), 144-152.
- Malach, R., Levy, I., Hasson, U., 2002. The topography of high-order human object areas. *Trends in Cognitive Sciences*, 6(4), 176-184.
- Marcas, V. L., Loenneker, T., Straessle, A., Jaggy, S., Kucian, K., Martin, E., 2004. An fMRI study of the cerebral macro network involved in 'cue invariant' form perception and how it is influenced by stimulus complexity. *Neuroimage*, 23(3), 947-955.
- Matsuo, K., Chen, S. H. A., Hue, C. W., Wu, C. Y., Bagarinao, E., Tseng, W. Y. I., Nakai, T., 2010. Neural substrates of phonological selection for Japanese character Kanji based on fMRI investigations. *Neuroimage*, 50(3), 1280-1291.
- Matthew Brett, Jean-Luc Anton, Romain Valabregue, Jean-Baptiste Poline. Region of interest analysis using an SPM toolbox [abstract] Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. Available on CD-ROM in NeuroImage, Vol 16, No 2.
- McCandliss, B. D., Cohen, L., Dehaene, S., 2003a. The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, 7(7), 293-299.
- McCandliss, B. D., Cohen, L., Dehaene, S., 2003b. The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn Sci*, 7(7), 293-299.
- Murray, S. O., Wojciulik, E., 2004. Attention increases neural selectivity in the human lateral occipital complex. *Nature Neuroscience*, 7(1), 70-74.

- Op de Beeck, H. P., Haushofer, J., Kanwisher, N. G., 2008. Interpreting fMRI data: maps, modules and dimensions. *Nat Rev Neurosci*, 9(2), 123-135.
- Petersson, K. M., Reis, A., Ingvar, M., 2001. Cognitive processing in literate and illiterate subjects: A review of some recent behavioral and functional neuroimaging data. *Scandinavian Journal of Psychology*, 42(3), 251-267.
- Petersson, K. M., Silva, C., Castro-Caldas, A., Ingvar, M., Reis, A., 2007. Literacy: a cultural influence on functional left-right differences in the inferior parietal cortex. *European Journal of Neuroscience*, 26(3), 791-799.
- Price, C. J., 2000. The anatomy of language: contributions from functional neuroimaging. *Journal of Anatomy*, 197, 335-359.
- Price, C. J., 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, 62(2), 816-847.
- Price, C. J., Devlin, J. T., 2003a. The myth of the visual word form area. *Neuroimage*, 19(3), 473-481.
- Price, C. J., Devlin, J. T., 2003b. The myth of the visual word form area. *Neuroimage*, 19(3), 473-481.
- Price, C. J., Devlin, J. T., 2011. The Interactive Account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, 15(6), 246-253.
- Reddy, L., Moradi, F., Koch, C., 2007. Top-down biases win against focal attention in the fusiform face area. *Neuroimage*, 38(4), 730-739.
- Reich, L., Szwed, M., Cohen, L., Amedi, A., 2011. A Ventral Visual Stream Reading Center Independent of Visual Experience. *Current Biology*, 21(5), 363-368.
- Reinke, K., Fernandes, M., Schwindt, G., O'Craven, K., Grady, C. L., 2008. Functional specificity of the visual word form area: General activation for words and symbols but specific network activation for words. *Brain and Language*, 104(2), 180-189.
- Roberts, D. J., Woollams, A. M., Kim, E., Beeson, P. M., Rapcsak, S. Z., Lambon Ralph, M. A., 2012. Efficient Visual Object and Word Recognition Relies on High Spatial Frequency Coding in the Left Posterior Fusiform Gyrus: Evidence from a Case-Series of Patients with Ventral Occipito-Temporal Cortex Damage. *Cereb Cortex*.
- Rolls, E. T., Stringer, S. M., 2006. Invariant visual object recognition: A model, with lighting invariance. *Journal of Physiology-Paris*, 100(1-3), 43-62.
- Song, Y. Y., Hu, S. Y., Li, X. T., Li, W., Liu, J., 2010. The Role of Top-Down Task Context in Learning to Perceive Objects. *Journal of Neuroscience*, 30(29), 9869-9876.
- Starrfelt, R., Gerlach, C., 2007. The visual what for area: Words and pictures in the left fusiform gyrus. *Neuroimage*, 35(1), 334-342.
- Starrfelt, R., Habekost, T., Leff, A. P., 2009. Too Little, Too Late: Reduced Visual Span and

- Speed Characterize Pure Alexia. *Cerebral Cortex*, 19(12), 2880-2890.
- Striem-Amit, E., Cohen, L., Dehaene, S., Amedi, A., 2012. Reading with Sounds: Sensory Substitution Selectively Activates the Visual Word Form Area in the Blind. *Neuron*, 76(3), 640-652.
- Tallon-Baudry, C., Bertrand, O., Henaff, M. A., Isnard, J., Fischer, C., 2005. Attention modulates gamma-band oscillations differently in the human lateral occipital cortex and fusiform gyrus. *Cerebral Cortex*, 15(5), 654-662.
- Tarr, M. J., Gauthier, I., 2000. FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, 3(8), 764-769.
- Turkeltaub, P. E., Gareau, L., Flowers, D. L., Zeffiro, T. A., Eden, G. F., 2003. Development of neural mechanisms for reading. *Nature Neuroscience*, 6(7), 767-773.
- Twomey, T., Duncan, K. J. K., Price, C. J., Devlin, J. T., 2011. Top-down modulation of ventral occipito-temporal responses during visual word recognition. *Neuroimage*, 55(3), 1242-1251.
- Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., Cohen, L., 2007. Hierarchical coding of letter strings in the ventral stream: Dissecting the inner organization of the visual word-form system. *Neuron*, 55(1), 143-156.
- Wandell, B. A., Rauschecker, A. M., Yeatman, J. D., 2012. Learning to See Words. *Annual Review of Psychology*, Vol 63, 63, 31-53.
- Warrington, E. K., Shallice, T., 1980. Word-Form Dyslexia. *Brain*, 103(Mar), 99-112.
- Wendelken, C., Baym, C. L., Gazzaley, A., Bunge, S. A., 2011. Neural indices of improved attentional modulation over middle childhood. *Developmental Cognitive Neuroscience*, 1(2), 175-186.
- Wojciulik, E., Kanwisher, N., Driver, J., 1998. Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. *Journal of Neurophysiology*, 79(3), 1574-1578.
- Wu, J. L., Li, X. J., Yang, J. J., Cai, C., Sun, H. Z., Guo, Q. Y., 2012. Prominent activation of the bilateral inferior parietal lobule of literate compared with illiterate subjects during Chinese logographic processing. *Experimental Brain Research*, 219(3), 327-337.
- Xu, G. F., Jiang, Y., Ma, L. F., Yang, Z., Weng, X. C., 2012. Similar spatial patterns of neural coding of category selectivity in FFA and VWFA under different attention conditions. *Neuropsychologia*, 50(5), 862-868.
- Xu, Y., 2005. Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex*, 15(8), 1234-1242.
- Yeatman, J. D., Rauschecker, A. M., Wandell, B. A., 2013. Anatomy of the visual word form area: Adjacent cortical circuits and long-range white matter connections. *Brain and*

Language, 125(2), 146-155.

Yoncheva, Y. N., Zevin, J. D., Maurer, U., McCandliss, B. D., 2010. Auditory Selective Attention to Speech Modulates Activity in the Visual Word Form Area. *Cerebral Cortex*, 20(3), 622-632.

Publications

1. Different Attentional Modulation in the Visual Word Form Area and Parahippocampal Place Area

Geqi Qi, Bin Wang, Jinglong Wu, Satoshi Takahashi, Seiichiro Ohno, Susumu Kanazawa
Neuroscience and Biomedical Engineering, in press (2014)

2. Similar Ventral Occipito-Temporal Cortex Activations in Literate and Illiterate Adults during the Chinese Character Matching Task: An fMRI study

Geqi Qi, Xiujun Li., Tianyi Yan, Bin Wang, Jiajia Yang, Jinglong Wu, Qiyong Guo
Neuroscience Letters, in press (2014)

3. Different occipito-temporal activations in Chinese literate and illiterate subjects during Chinese character processing

Geqi Qi, Xiujun Li, Tianyi Yan, Jinglong Wu, Qiyong Guo
Complex Medical Engineering (CME), 2010 IEEE/ICME International Conference Proceeding
pp.134-137

4. Differential sensitivity for house pictographs and Chinese logographs in the “parahippocampal place area”

Geqi Qi, Bin Wang, Takahashi, S., Ohno, S., Kanazawa, S., Jinglong Wu
Complex Medical Engineering (CME), 2013 IEEE/ICME International Conference Proceeding
pp.444 – 447

5. Functional Role of the Left Ventral Occipito-Temporal Cortex in Reading

Geqi Qi, Jinglong Wu
Biomedical Engineering and Cognitive Neuroscience for Healthcare: Interdisciplinary Applications, ed. Jinglong Wu (2013)
pp.192-200

Acknowledgements

I would like to express the deepest appreciation to my committee chair Professor Jinglong Wu, who has the attitude and the substance of a genius: he continually and convincingly conveyed a spirit of adventure in regard to research and scholarship, and an excitement in regard to teaching. Without his guidance and persistent help this dissertation would not have been possible.

I would also like to thank my committee members, Professor Shinichiro Yanase and Professor Akihiko Horibe for serving as my committee members even at hardship. I also want to thank you for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions, thanks to you.

Special thanks to Associate Professor Satoshi Takahashi for helping me in many ways from research to general affairs since the first day I arrived in Japan. Furthermore, a big thank you to Assistant Professor Jiajia Yang, who have instructed and helped me with experiments, analysis and writing paper even with tight schedule.

I would especially like to thank all the people who were involved in my study: the participants and the staff, especially Professor Qiyong Guo of Shengjing hospital in Shenyang, China and Professor Susumu Kanazawa, Dr. Seiichiro Ohno in Okayama University hospital in Japan. I also thank the staff and students of Biomedical Engineering Laboratory, Okayama University. Without their supports and help, the study would not be completed.

Finally, I would like to express my greatest gratitude to my parents for always providing me with love and support needed in whatever I made my mind up to do. With them backing me up, I have been able to continually push myself through difficult times.