3D shape and orientation representation in the human cortical areas

by

Zhen LI

Student ID Number: 1208009

A dissertation submitted to the
Engineering Course, Department of Engineering,
Graduate School of Engineering,
Kochi University of Technology,
Kochi, Japan

For the degree of
Doctor of Engineering

Assessment Committee:
Supervisor: Associate Prof. Hiroaki SHIGEMASU, School of Information
Co-Supervisor: Prof. Keizo SHINOMORI, School of Information
Co-Supervisor: Associate Prof. Hiroshi KADOTA, School of Information
Committee Member: Prof. Kiyoshi NAKAHARA, School of Information
Committee Member: Associate Prof. Yukinobu HOSHINO, School of Systems Engineering

September 2019

ABSTRACT

3D shape and orientation representation in the human cortical areas

Zhen LI
Doctoral Program of Engineering Course, Department of Engineering,
Graduate School of Engineering

The ability of human to perceive three-dimensional (3D) shape and orientation of the real world is vital in daily life. There are a variety of depth cues that can be used to extract 3D perception by human brain: binocular disparity, perspective, motion parallax, texture and so on. Although there are a lot of studies about how brain extract information from these cues to perceive 3D shape and orientation, the underlying neural mechanisms have not been fully investigated. To this end, we adopted functional magnetic resonance imaging (fMRI) method which can measure blood oxygenation level-dependent (BOLD) signal of the brain. Multi-voxel pattern analysis (MVPA) method was used to judge whether the activity pattern of neurons reflects a particular feature of stimuli. Furthermore, by transfer classification in which we trained a classifier on one type of stimuli and test the classifier on another type of stimuli, we investigated whether common representation of different type of stimuli is involved in regions of interest (ROIs). ROIs including early visual areas which were defined by standard retinotopic mapping procedure and higher visual areas which were defined by standard localizers. We focused on two kinds of depth cues: binocular disparity and perspective. Two topics were investigated:

In the first topic, binocular disparity cue was investigated. We investigated convex–concave shape representation and horizontal–vertical orientation representation of stereoscopic surface. Stimuli were curved convex–concave surfaces in horizontal or vertical orientation. Surfaces were simulated in two depth positions. Participants viewed the stimuli while BOLD signal was measured. A series of classification was performed. (1) to investigate convex–concave shape representation among ROIs, in general, two types of convex–concave shape classification were performed: same-type stimuli convex–concave classification which was used to verify whether neurons are selective to convex–concave stereoscopic surface shapes in ROIs, and transfer convex–concave classification on surfaces in different orientation and depth position to verify whether generalized representation of shapes is involved in ROIs. Results showed that neurons in most of the ROIs are selective to convex–concave

stereoscopic surface shapes, and more important, V3A showed more generalized representation of shapes irrespective orientation and depth position of surfaces. (2) similarly, to investigate horizontal–vertical orientation representation of stereoscopic surface, two types of horizontal–vertical classification were performed: same type stimuli horizontal–vertical classification which was used to verify whether neurons are selective to horizontal–vertical orientation of stereoscopic surface in ROIs, and transfer horizontal–vertical classification on surfaces of different shape at same depth position which was used to verify whether generalized representation of orientations is involved in ROIs. Results indicated that neurons in all ROIs are selective to orientation of stereoscopic surface and higher dorsal areas V3A and V7, ventral areas LOC and intraparietal sulcus areas VIPS and POIPS are involved in more generalized representation of orientation irrespective of shape of surfaces.

In the second topic, binocular disparity cue and perspective cue were investigated. The main purpose of this topic is to investigate whether ROIs are involved in representation of convex–concave 3D shapes from binocular disparity or perspective and investigate whether common representation of shapes from these two different cues is involved in ROIs. Stimuli were convex–concave shape consisted of two slanted planes which were defined by binocular disparity or perspective respectively. In detail, three types of stimuli were used: shapes defined by RDS, shapes defined by black–white dotted lines with perspective and shapes defined by black–white dotted lines with disparity. Two different disparity stimuli types (RDS and black-white dotted lines with disparity) were adopted to verify whether shapes from disparity but different elements share common representation. Two main types of classification were performed: (I) Same cue type stimuli convex–concave classification. In this type of classification, the SVM was trained and tested using data of the same cue type. Corresponding to the three types of stimuli, this type of classification includes three subtypes of classification: same cue type stimuli convex–concave classification using data of RDS, same cue type stimuli convex–concave classification using data of lines with perspective, and same cue type stimuli convex–concave classification using data of lines with disparity. The purpose of this type of classification was to verify whether neurons in ROIs are selective to convex–concave shape defined by one of these three types of stimuli. (II) Transfer convex–concave classification using stimuli of different cues. In this type of classification, convex–concave classifications were performed between combinations of different types of stimuli. Corresponding to these three types of stimuli, this type of classification also includes three sub classification: transfer convex–concave classification between data of RDS and lines with perspective, transfer convex–concave classification between data of lines with perspective and lines with disparity, and transfer convex–concave classification between data of RDS and lines with disparity. The purpose of this type of classification is to investigate whether a

common neural activity pattern is involved for each ROI in the processing of shape from the two different types of stimuli. Results showed that early and middle visual areas had a tendency to be high classification accuracy in the "same cue type stimuli convex–concave classification," and for some higher visual areas, in particular the area dorsal intraparietal sulcus (DIPS), classification accuracy had a tendency to be high in both "same cue type stimuli convex–concave classification" and "transfer convex–concave classification using stimuli of different cues.

In summary, our research investigated generalized representation of convex–concave shape and horizontal–vertical orientation of curved surfaces defined by binocular disparity, and we also investigated common representation convex–concave representation of shapes which consisted of slanted planes defined by binocular disparity or perspective, respectively. This dissertation provides a better understanding of the neural representation of shape and orientation of 3D object among human cortical areas.

Keywords: shape perception, orientation perception, fMRI, MVPA, depth cues

Table of contents

List of Figures

CHAPTER 1

INTRODUCTION

Brain's ability to perceive three-dimensional (3D) world is vital for the human survival. Figure 1.1 shows the front view from a car running on the road. The driver need to judge the width of the road, the distance to the other cars and the position of obstacles. These tasks highly depend on 3D perception.

Figure 1.1 The front view from a car running on the road. This picture was taken on the great seto bridge in Japan.

The images projected onto the retina of our eyes are two dimensional (2D), how can the human brain actual achieve 3D is an intriguing question. The main purpose of my research is to investigate the neural mechanism for how our brain perceives 3D in cortical areas. This chapter is about general knowledge related to human 3D perception. In section 1.1, we will first introduce the visual system from the eyes to the visual cortex which is the foundation of human vision. Then, we will introduce 3D depth cues which can be used to infer 3D information by the eyes. Section 1.2 is about literature

review and motivation of our study. Finally, in section 1.3, we will describe the overview of this dissertation.

1.1 Human 3D vision

1.1.1 Visual system

When the light of outside world goes into the eyes, it will initiate a series of neural activities in the visual system, and after that, we will perceive the image of the world. A series of optical and neural transformation are required to form visual perception. The cornea and lens transformed the light which arriving at the eyes, and a retinal image was created. After that, the photoreceptors, which is the light-sensitive elements of the eyes, transform the retinal image into neural responses. These neural responses are then transformed into neural representations in the optic serve, and finally transformed into representations among cortical areas. In general, the visual system is composed of three main parts: the eye, the visual pathway and the visual cortex.

(1) The eye

Figure 1.2 shows a cross-section of the human eye. By receiving photons and guide them onto photoreceptors, the eye starts the process of vision. It is a light-tight and roughly spherical chamber. In the inside of surface, it has a sheet of photoreceptors. There is an opening in the chamber which covered by a transparent membrane, and this opening allows light to pass through. The transparent membrane is termed the cornea. After passing through the cornea, the incoming light pass through an aperture termed as the pupil. The pupil is formed by a muscular diaphragm known as the iris. The light pass through the lens and arrive at the retina which is located behind the pupil.

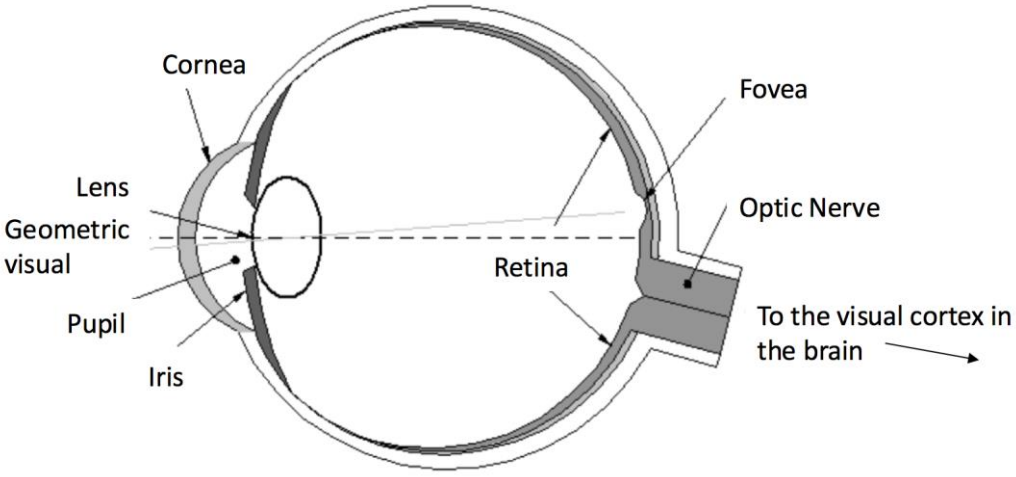

Figure 1.2 Schematic diagram of the cross-section through a human eye. Image from Fritzsche et al. (2010).

The retina, which is located near the optic nerve, is a thin tissue layer on the back of the eye on the inside. The main function of the retina is to accept the light that the lens has focused, convert it into neural signals, and after that these signals were sent to visual cortex of the brain serving for visual perception. The retina is a layered network and there are five types of cells which are contained in it. The nuclei of the cells are grouped into three layers: in the outer layer, there are photoreceptor cells, which can transform the light signal into neural signals; in the ganglion cell layer, the ganglion cell fibres carry neural signals out of the eyes and along the visual pathway to the brain; in the nuclear layer, the complex retinal circuitry connects the photoreceptors to ganglion cells.

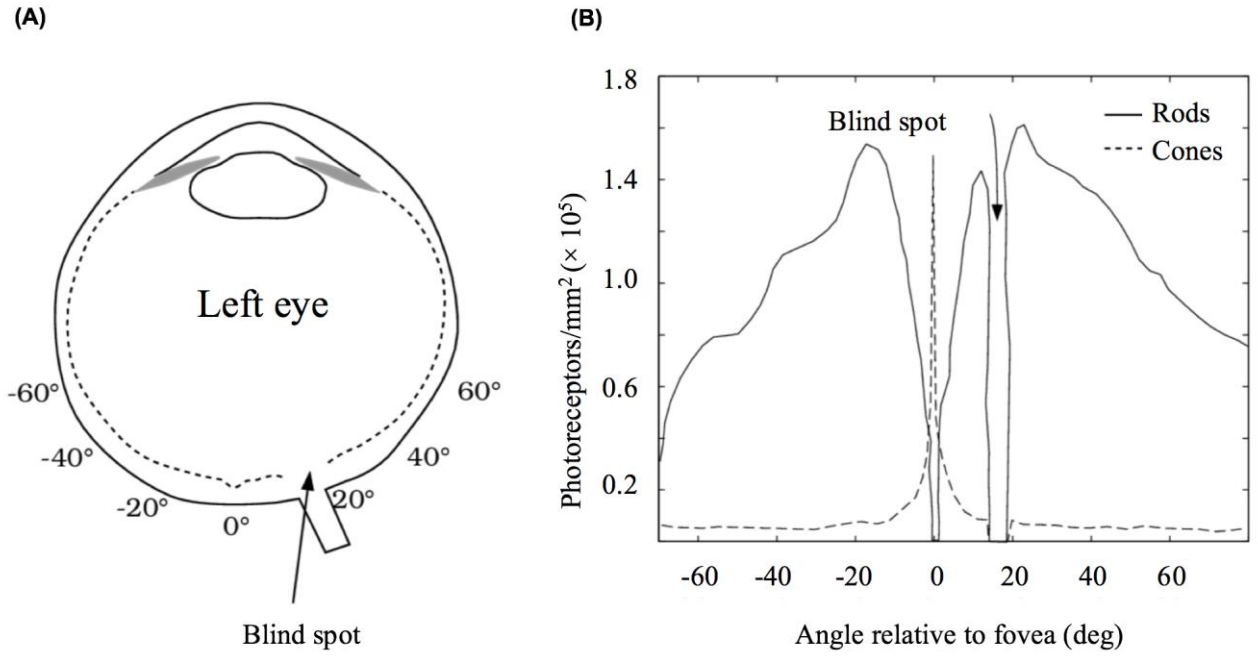

Figure 1.3 The density of cone and rod photoreceptors in the human retina. (A) Illustration of visual angle relative to the foveal position of the left eye; also the position of blind spot is shown. (B) The cone receptors are dense in foveal and the density decrease shapely toward the peripheral region. The density of rod photoreceptors is low in the fovea, it increases toward the peripheral region, and reach the highest density between 10 and 20 degree peripheral to the fovea. There are no photoreceptors in the blind spot. Image from Wandell (1995).

There are more than 100 million photoreceptor cells on the retina of each eye. The photoreceptor is a thin and long tube. The outer segment of it contains light-sensitive pigment and the inner segment in effect forms the cell body. According to their shapes, the photoreceptors are separated into two types: termed the rods and cones in the human eye. In general, the outer segments of the rods are rod-shaped, and the outer segments of cones are cone-shaped. For each eye, there are about 100 million rods and 5 million cones. The rods and cones contained pigments with different sensibility to light: the pigment of rods is very sensitive to light at low levels, whereas the pigment of cones is very

sensitive to light at high level. The rods and cones are not distributed on the retina evenly. The region of highest visual acuity is in a central area of retina known as fovea (Figure 1.2). Figure 1.3 shows the distribution of rods and cones photoreceptors. The cones are mainly located in the fovea and there are about 50,000 cones in the central human fovea. There are no rods found in the fovea. The ganglion cell axons exit the retina at the optic disk. There are no photoreceptors in this spot, and this is a blind spot in the retina.

(2) The visual pathway

The optic nerve is formed by the fibres of the retinal ganglion cells. At the blind spot, these fibres leave the eye and start their trips to the cortex, and finally reach to six locations within the brain. Most of the projection is destined to lateral geniculate nuclei (LGN), and about 90% of ganglion cell fibres are received by LGN. The remaining 10% is shared among the other sub-cortical nuclei, though most goes to the superior colliculi (Perry et al. 1984). It should not be underestimated the importance of the minor projections. The 10% of optic nerve fibres still exceeds the total central projections from taste, smell, and hearing.

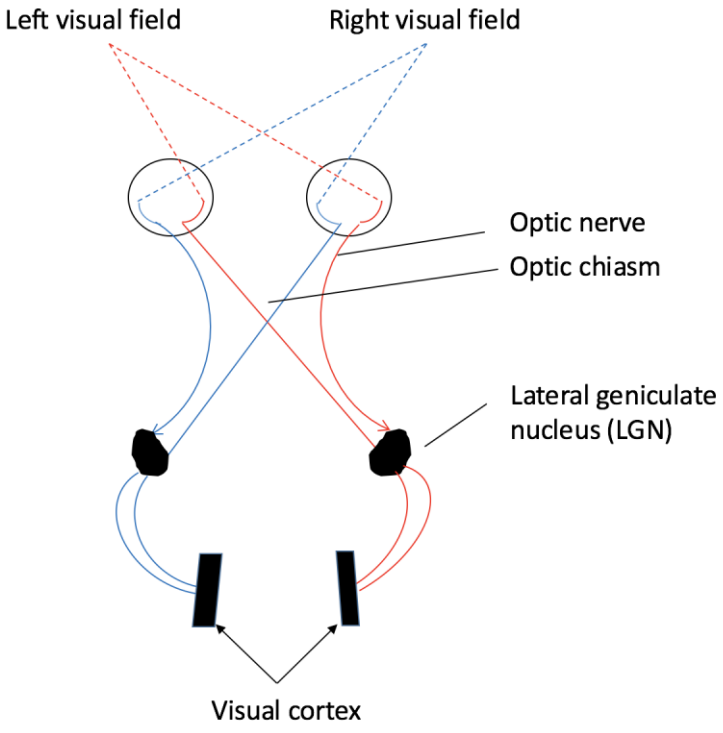

Figure 1.4 The visual pathway from the retina to the visual cortex

Figure 1.4 shows the main visual pathway from the retina to the visual cortex. The optic nerves which start from the two eyes meet at a place termed optic chiasm. In this place, a partial decussation happens which means that a proportion of these fibres cross over the other side of the brain while the

other fibres do not. The axons from ganglion cells on the nasal side of each retina cross to the other side, and those on the temporal size do not cross to the other size. As a result, at this stage, the nerve fibres on the left side contain information of the right visual field, that is the right part of the world; and the nerve fibres on the right side contain information of the left visual field, that is the left part of the world. Therefore, although the left LGN and the cortex to which it projects receive information from both of the eyes, they contain information only about the right half of the visual field. Similarly, although the right LGN and the cortex to which it projects receive information from both of the eyes, they contain information only about the left half of the visual field.

(3) Visual Cortex

The human cortex is a sheet of neurons with a thick about 2 mm and a surface area about 1400 cm^2. Unlike the retina lines the eye, the visual cortex does not line the skull, it is like a crumpled sheet stuffed into the skull. A gyrus is a ridge which formed by the folded cortex and can be seen from the exterior, while a sulcus is a shallow furrow that separates a pair of gyri. The human brain is divided into four lobes which are marked by the most visible sulci. These lobes include: frontal lobe, parietal lobe, temporal lobe, and occipital lobe. Figure 1.5 shows the positions of these lobes. For each lobe, these are many distinct brain areas, that is there are many groups of continuous cortical neurons that may work in an interrelated way.

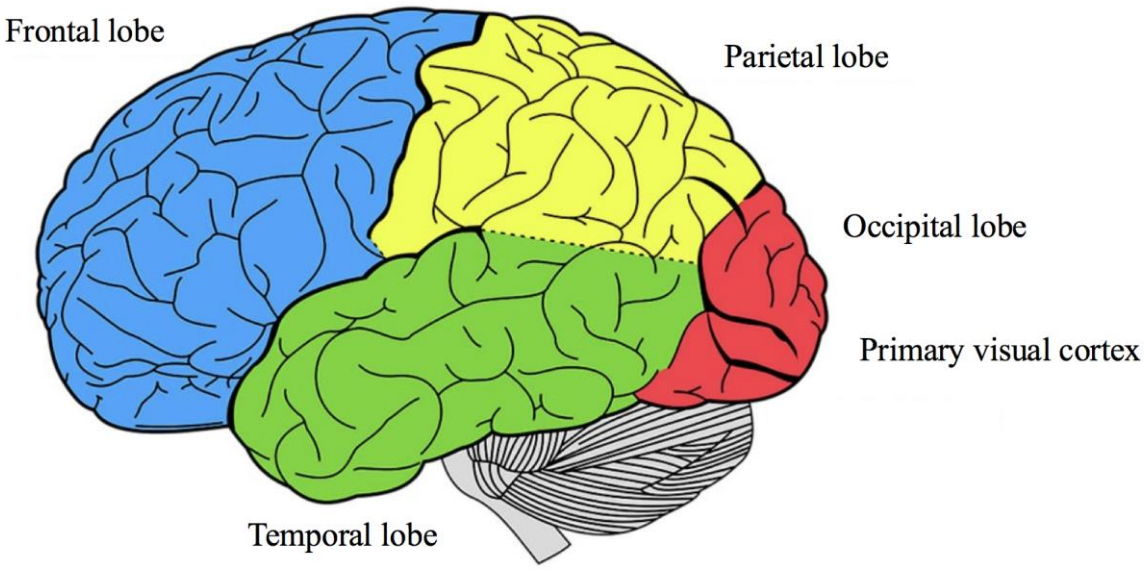

Figure 1.5 Cortical regions of human.

The majority of visual information from the retina and the LGN gets to an area in the occipital lobe named area V1 (also known as the primary visual cortex). All the other visual areas (also known as

extrastriate cortex) are highly dependent on the output of V1, directly or indirectly. The V1 is a large cortical area containing about 1.5×10^8 neurons, which is much larger than the number of neurons in the LGN, which is about 10^6. The V1 area can be divided into 6 layers, which number from 1 to 6 from the layer nearest to the surface and the layer deepest. Furthermore, layer 4 can be divided into separate sub-layers name 4A, 4B, 4Cα, and 4Cβ. The functions of these layers are different: some layers directly receive information the LGN, whereas other layers do not. Some layers act as the direct or indirect input of higher visual areas, and in fact, the layer 6 act as a source of the input of LGN which accept information back to LGN.

Beside early visual area V1, there are many visual cortical areas have been identified. These areas are identified by different methods, including topographic organization, anatomical connections, and response properties of neurons. Brain-imaging method has also been used to identify visual areas (Tootell et al., 1998). It is still not fully understanding the functional importance of the diversity of areas. In order to reveal how visual processing advances from one area to another, researchers have tried to organize these visual areas in a hierarchical way based on the anatomical connections between them (Felleman et al., 1991). However, as there are about 300 projections between different visual areas and the possible combination of connections between areas is large, it is very different to get a final conclusion about the hierarchy between visual areas.

1.1.2 Depth cues for 3D vision

The images casted onto the retina of human eyes are 2D, but our brain can perceive the world in a 3D way. Depth perception is the ability to see the world in a 3D way and to judge distance of objects. Our visual system need to recover the 3D information from the 2D images. The clues that can be used for recovering the 3D information are termed depth cues. The are many visual cues on the retinal images which can be used to infer the depth of the objects in the view. In addition, the non-visual cues from oculomotor muscles controlling eye position and the accommodative state can be used by human brain to infer depth position. The depth cues can be classified into two types: binocular cues are cues depend on information taken by two eyes, whereas, monocular cues are cues that depend on information taken by one eye. Binocular cues include binocular disparity, vergence and so on; monocular cues include perspective, texture gradients, motion parallax, retinal image size, interposition and so on.

1.1.2.1 Binocular cues

(1) Binocular disparity

In human, the eyes are typically 5.0 - 7.5 cm apart in horizontal orientation. Therefore, the retinal image of each eye are slightly different. This slightly differences provide one of the most powerful depth cues. The depth position of an object can be calculated with a high accuracy by using the different images. For the stereo disparity systems in the brain, there are two types of disparities: absolute disparity and relative disparity. Imagining that a person is observing a point P in the space (Figure 1.6), then the image of P falls on the foveal of each eye. For another point Q in a different depth position, the image of Q is located on the different positions of the left and right retinae relative to the foveae. In order to observe Q directly, the two eyes need to rotate with different angles. The difference between these two angles is termed as absolute disparity of Q. Relative disparity is another way to measure the relationship between point P and Q. The relative disparity is not related to where the eyes are fixating on. Suppose the angular separation of P and Q seen by the left eye is α, and the separation seen by the right eye is β, α is larger than β, then the difference between angular separations (α - β) is called the relative disparity between point P and Q. Two simultaneously visible targets are required to defined relative disparity. The relative disparity is a sensitive way to measure spatial relationship between two visual targets in the three-dimensional space.

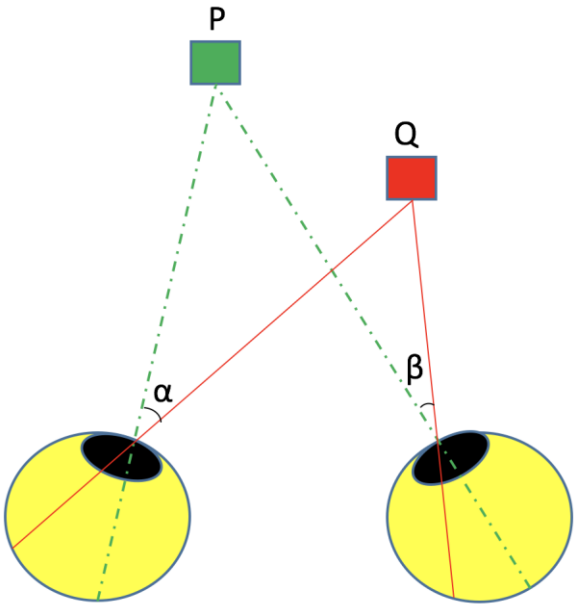

Figure 1.6 The geometrical basis of binocular disparity.

(2) Vergence

When two eyes fixate on the same point in space, their visual axes need to converge so that the

image of the point can be projected onto the fovea in each eye. The vergence angle, which is formed by the intersection of visual axes, is dependent on the distance of the point in space. Vergence angel is controlled by the extra-ocular muscles, it offers a cue about the absolute distance of the fixated point. Figure 1.7 shows the relationship between the distance and vergence angle. From the figure, we can see that the further the object is, the smaller the vergence angle is; and the nearer the object is the larger the vergence angle is.

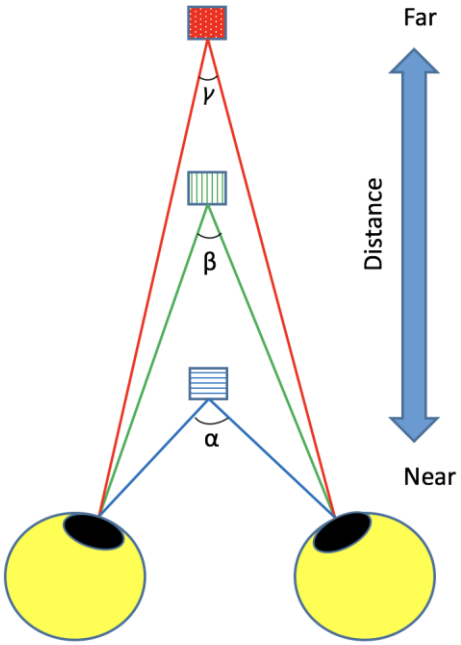

Figure 1.7 The relationship between distance and vergence angle.

1.1.2.2 Monocular cues

(1) Linear perspective

Linear perspective is a depth cue that is related to both relative size and texture gradient. In linear perspective cue, as parallel lines extend far away, these liens seem to get more and more closer and converge in the distance. The more the lines converge, the further away they appear. The feature allows the brain to perceive the view in a 3D way. Consider an example that we are looking at parallel railways which extend to the distance (Figure 1.8), they appear to diminish towards a single point far ahead of you.

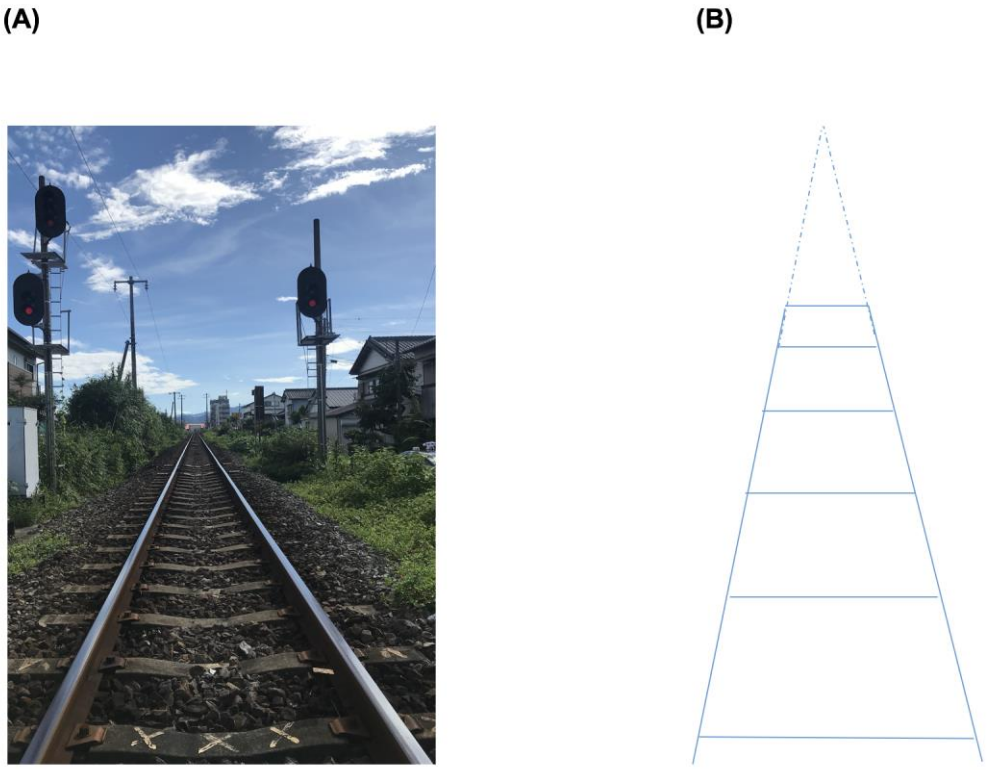

Figure 1.8 Linear perspective depth cue. (1) Example of depth from linear perspective (the photo was taken near tosayamada station in Japan). (2) The geometrical basis of linear perspective.

(2) Texture

Texture is a monocular cue provided by surface with texture. The size, shape, and density of texture elements variate depending on the distance and shape of object. This provide an important information for human brain to perceive 3D. Figure 1.9 shows examples of shape depicted by texture.

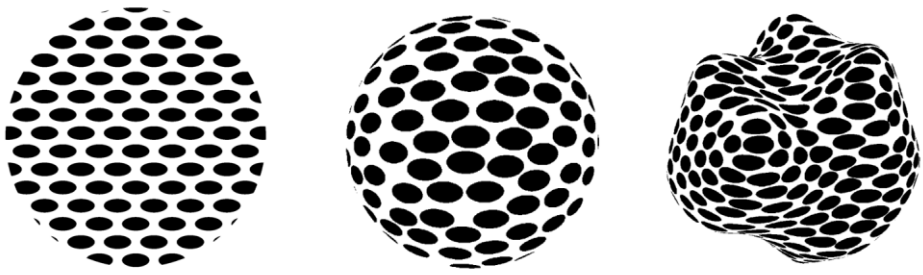

Figure 1.9 3D shapes from texture. Image from Todd (2010).

(3) Motion Parallax

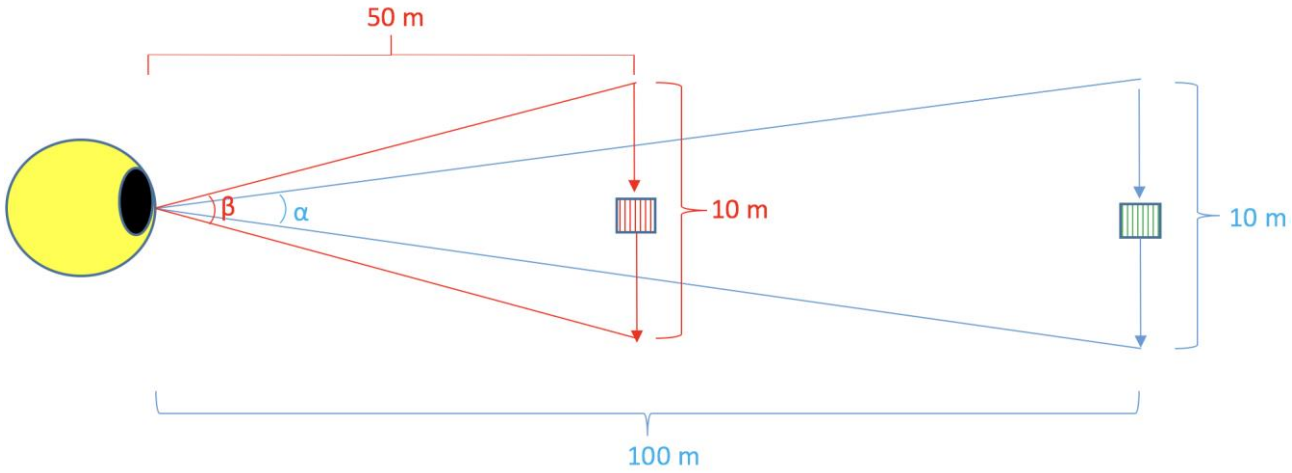

Figure 1.10 Illustration of relationship between angle and distance object moves.

In previous part, the depth cues discussed are all static cues, whereas motion parallax is a cue when the objects are moving or the observer is moving relative to the objects. In this kind of cue, objects that are closer to observer appear to move faster than objects that are farther. It is a kind of monocular cue which can be perceived through one eye. The reason for this phenomenon is related to the angle the object moves compared to the distance the object moves. Figure 1.10 shows an object move 10 meters in two different distances of 100 meters and 50 meters from the eyes. The corresponding angle $\alpha \approx 0.1$ degree and $\beta \approx 0.2$ degree. Obviously, the angle in the near distance is larger than the that of in far distance.

1.2 Literature review and motivation

The interactions with the outside world often require 3D spatial shape and orientation information. The visual images projected onto the retina of two eyes were 2D, our brain need to reconstruct the 3D information of the world from the 2D images. A variety of depth cues can be used by human for reconstructing the 3D image, including binocular disparity, perspective, texture, and so on. In order to understand how our brain processing the visual information and how to construct the 3D, many studies were conducted. Here we briefly summarize the related studies.

One of the most important theory in the visual information processing is “two-stream” theory, which was originally proposed by Mishkin and Ungerleider (1982) based on lesion studies of non-human primates. In this theory, visual information is processed progressively with each area processing information based on the output of previous areas, and after visual information processed in V1, the

visual processing stream is spitted into two anatomical and functional different streams, with the dorsal from dorsal area V1 to parietal cortex areas, and ventral stream from V1 to temporal areas. The dorsal stream mainly processes information about an object's location and spatial relationships and has been termed as the "where" stream, and the ventral stream mainly processes information about an object's shape and identity and has been termed as the "what" stream. Later, this theory was revised by Goodale and Milner (1992), instead, they argued that the ventral stream serves for conscious vision (known as the "perception" stream) and the dorsal stream serves for the unconscious control of action, such as the movement of the body guided by visual input (known as the "action" stream). There are lots of studies support the "dorsal action" and "ventral perception" separation of two streams, including single-unit recording (Sakata et al., 1995; Fogassi et al., 2001), neuropsychological studies (Goodale et al., 1991; James et al., 2003), behavioral psychophysics (Ganel et al., 2003), and functional imaging (Culham et al., 2003; Monaco et al., 2014), and among others. However, there is no complete agreement on how visual information is processed on the streams.

As to the 3D information processing, areas that are selective to 3D shape are widely distributed among visual cortex, including early visual areas (V1-V3) where basic features such as binocular disparity, texture, luminance are detected locally, middle ventral visual areas such as V4, and areas further along ventral stream such as lateral occipital cortex (LOC) and ventral temporal cortex (Grill-Spector and Weiner, 2014; Janssen P et al., 2000; Moore and Engel, 2001). Dorsal cortex is likely to be related to high order properties about 3D object structure, such as slant and curvature (Erlikhman et al., 2018). As the 3D information from different cues processed along the visual pathways, these depth cues concurrently contribute to construct 3D object. It is possible in some area, depth information from different cues is fused and a final depth estimator about each point of the object is generated. Therefore, a generalized representation of 3D object may be existing, and this representation is not relying on depth cue type directly. Meanwhile, this representation should not rely on local features of stimuli. That is, a common neural activity pattern related 3D objects may exist in some area irrespective cue type or local feature of stimuli. In order to investigate the whether some area contain common neural activity pattern which related to 3D object, we adopted multi-voxel pattern analysis (MVPA), specifically, transfer classification which trained classifier on fMRI data with stimuli of one type and tested classifier on fMRI data with stimuli of another type was used. This method can be used to verify whether a common activity pattern of neurons is involved during processing different types of stimuli. In detail, two topics were investigated. In the first topic, binocular disparity cue was investigated. Stimuli defined by disparity with same shape but different

orientation were used. In the second topic, depth cues binocular disparity and perspective were investigated. We investigated whether a common activity pattern exist during the processing of shapes from different cues.

1.3 Overview of our research

Topic 1:

- Investigate selectivity to shape of surface defined by disparity on each ROI. Whether the selectivity is dependent on disparity directly or dependent on generalized representation of shape?
- Investigate selectivity to orientation of surface defined by disparity on each ROI. Whether the selectivity is dependent on disparity directly or dependent on generalized orientation representation?

Topic 2:

- Whether activity patterns evoked by shape defined by one type of stimuli (disparity or perspective) can afford classification of patterns evoked by shape defined by another type of stimuli (perspective or disparity respectively)?

1.4 References of CHAPTER 1

Culham JC, Danckert SL, DeSouza JF, Gati JS, Menon RS, and Goodale MA (2003). Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. Exp. Brain Res. 153, 180-189.

Erlikhman G, Caplovitz GP, Gurariy G, Medina J, and Snow JC (2018). Towards a unified perspective of object shape and motion processing in human dorsal cortex. Consciousness and Cognition 64: 106-120.

Felleman DJ, and Van Essen DC (1991). Distributed hierarchical processing in the primate cerebral cortex. Cerebral cortex, 1, 1-47.

Fogassi L, Gallese V, Buccino G, Craighero L, Fadiga L, and Rizzolatti G (2001). Cortical mechanism for the visual guidance of hand grasping movements in the monkey: A reversible

inactivation study. Brain 124, 571-586.

Fritzsche H, Poirier E, Haagsma J, Ophus C, Luber E, Harrower CT, and Mitlin D (2010). A systematic neutron reflectometry study on hydrogen absorption in thin $Mg_{1-x}Al_x$ alloy films. Can J Phys. 88, 723-728.

Ganel T, and Goodale MA (2003). Visual control of action but not perception requires analytical processing of object shape. Nature 426, 664-667.

Grill-Spector K, and Weiner KS (2014). The functional architecture of the ventral temporal cortex and its role in categorization. Nat Rev Neurosci. 15, 536-548.

Goodale MA, Milner AD, Jakobson LS, and Carey DP (1991). A neurological dissociation between perceiving objects and grasping them. Nature 349, 154-156.

Goodale MA, and Milner AD (1992). Separate visual pathways for perception and action. Trends Neurosci. 15, 20-25.

James TW, Culham J, Humphrey GK, Milner AD, and Goodale MA (2003). Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study. Brain 126, 2463-2457.

Janssen P, Vogels R, and Orban GA (2000). Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. Sci. 288, 2054-2056.

Mishkin M, and Ungerleider LG (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Behav. Brain Res. 6, 57–77.

Monaco S, Chen Y, Medendorp WP, Crawford JD, Fiehler K, and Henriques DY (2014). Functional magnetic resonance imaging adaptation reveals the cortical networks for processing grasp-relevant object properties. Cereb. Cortex 24, 1540-1554.

Moore C, and Engel SA (2001). Neural response to perception of volume in the lateral occipital complex. Neuron 29, 277-286.

Perry VH, Oehler R, and Cowey A (1984). Retinal ganglion cells that project to the dorsal lateral geniculate nucleus in the macaque monkey. Neuroscience 12, 1101-1123.

Sakata H, Taira M, Murata A, and Mine S (1995). Neural mechanisms of visual guidance of hand action in the parietal cortex of the monkey. Cereb. Cortex 5, 429-438.

Todd JT (2010). The perception of 3D shape from texture based on directional width gradients. J Vis., 10, 1-13.

Tootell RBH, Hadjikhani NK, Mendola JD, Marrett S, and Dale AM (1998). From retinotopy to

recognition: fMRI in human visual cortex. Trends Cogn Sci. 2, 174-183.

Wandell BA (1995). Foundations of vision. Massachusetts: sinauer associates inc.

CHAPTER 2

TOPIC 1: Shape and orientation representation of stereopsis

2.1 Abstract

The brain's ability to extract 3D shape and orientation information from viewed objects is vital in daily life. Stereoscopic 3D surface perception relies on binocular disparity. Neurons selective to binocular disparity are widely distributed among visual areas, but the manner in these areas are involved in stereoscopic 3D surface representation is unclear. To address this, participants were instructed to observe RDS depicting convex and concave curved surfaces and the BOLD signal of visual cortices was recorded. Two surface types were: (i) horizontally positioned surfaces defined by shear disparity and (ii) vertically positioned surfaces defined by compression disparity. The surfaces were presented at different depth positions per trial. Functional magnetic resonance imaging data were classified from early visual areas to higher visual areas. We determined whether cortical areas were selective to shape and orientation by assessing same-type stimuli classification accuracies based on multi-voxel BOLD signal patterns per area. To identify whether some areas were related to a more generalized shape or orientation representation, transfer classification was used by training classifiers on one dataset type and testing classifiers on another type. Same-type stimuli classification results showed that most selected visual areas were selective to shape and all of them were selective to the orientation of disparity-defined 3D surfaces. Transfer classification results showed that in the dorsal visual area V3A, classification accuracies for the discriminate convex–concave shapes were higher than baseline of statistical significance for all types of classifications, demonstrating that V3A is related to generalized shape representation. Classification accuracies for discriminating horizontal–vertical surfaces in higher dorsal areas V3A and V7 and ventral area lateral occipital complex (LOC) as well as in some areas of intraparietal sulcus were higher than baseline of statistical significance, indicating their relation to the generalized representation of 3D surface orientation.

2.2 Introduction

The ability to interact with objects in the real world is closely related to 3D perception. This skill depends on at least two abilities as follows: (i) the ability to perceive the shape of a 3D object and (ii) the ability to judge the orientation of the object. For example, when someone attempts to pick up a pencil on a desk or insert a key into a lock, the procedure depends on the above-mentioned abilities. Although these activities are common and essential in human daily life, their underlying visual mechanisms have yet not been completely investigated. Binocular disparity, which is generated by the horizontal separation of the two eyes, is one of the most important cues for 3D perception. It is an extremely informative cue that is sufficient for depicting any 3D percept imaginable—the depth of the points of the object as well as the surface shape and orientation, which are considered higher-order surface properties.

To extract useful information from images registered by the two eyes, a sequence of processing stages is required (Marr & Poggio, 1976). Different brain regions may play different roles in 3D shape/orientation processing, with some areas responsible for low-level disparity and other areas responsible for middle or higher stage 3D representations that do not depend on low-level disparity.

Previous studies of 3D perception from disparity have focused on the neurons' selectivity to disparity depicting 3D objects. Zero-order depth indicates the depth position of an object. First-order depth corresponds to a linear gradient of depth, such as a plane slanted in depth. Second-order depth refers to curvature in depth, e.g., disparity curvature in stereoscopic processing (Orban et al., 2006b). In monkey and human brains, the neurons selective to binocular disparity exist at multiple levels of the visual hierarchy, starting from the early visual areas to the object-selective and motion-selective areas and parietal areas (Chandrasekaran et al., 2007; Cumming & DeAngelis, 2001; Neri, 2005; Parker, 2007). There is evidence that neurons tuned for 3D shape are widely distributed across the visual areas of the cortex. Janssen et al. (1999) verified that a portion of the inferotemporal (IT) neurons in macaques is selective for 3D shape depicted by binocular disparity. Janssen et al. (2001) conducted a single-cell method on animals and found that neurons of the lower bank of the superior temporal sulcus (TEs) can be selective for horizontal 3D shapes. Georgieva et al. (2009) tested the interaction between stereo and order of disparity, concluding that the V3A complex and certain intraparietal sulcus (IPS) regions can extract and process 3D shape from stereo. Recently, Alizadeh et al. (2018) studied a patch of the macaque posterior inferotemporal area TEO that is activated more by a curved surface than a flat surface. Using the single-cell method, they observed that this patch did not contain a large number of higher-order disparity-selective neurons. However, the sign of the

disparity gradient of the stimuli could reliably be classified using a linear SVM. Orientation discrimination research has primarily focused on two types of orientation as follows: (i) tilt, which refers to rotation in a 2D image plane, similar to a clock hand changing its orientation over time, and (ii) slant, which refers to rotation toward/away from the frontoparallel direction. In relation to selectivity to orientation, Nguyenkim and DeAngelis (2003) found that several middle temporal area (MT) neurons in rhesus monkeys are tuned to 3D surface orientation defined by binocular disparity, and that tilt and slant typically exhibit independent effects on MT responses. Rosenberg et al. (2013) reported that in the caudal intraparietal area (CIP) of macaques, an explicit representation of slant exists, indicating that this area plays an important role in encoding surface orientation information. In functional imaging studies of humans, Shikata et al. (2001) found that both posterior (CIP) and anterior (AIP) areas within the intraparietal sulcus responded to planes oriented in depth using texture gradients as monocular cues. In addition, Naganuma et al. (2005) found that orientation information defined by disparity is processed in the parietal area. Ban and Welchman (2015) revealed a dorsal hierarchy that extracts 3D surface orientation from binocular disparity, demonstrating that responses in the V3A judgments of slant in humans.

Previous studies for shape/orientation representation of 3D object focused mainly on these types of questions: (1) selectivity of neurons to specific depth cues (e.g., binocular disparity) using single-cell studies; (2) BOLD signal changes to specific stimuli (e.g., 3D shape) vs. control condition using fMRI studies; (3) distinguishing BOLD signal patterns caused by different shapes/orientations; and (4) integration of different cues that concurrently describe a 3D object. Our present study belongs to the third type. However, to the best of our knowledge, previous studies have not investigated whether the shape/orientation presentation of stereopsis is directly dependent on disparity or to a more generalized processing. We investigated whether a generalized representation is involved by transfer classification of BOLD signal patterns. In detail, we investigated the convex–concave shape representation and horizontal–vertical orientation representation of the surface across cortical visual areas. In the present study, we investigated the convex–concave shape representation and horizontal–vertical orientation representation of the surface across cortical visual areas. The main purpose of our study was to identify the areas containing a reliable representation of the shape and orientation of the 3D surface and to determine the extent of generalization of the representation. Depending on the degree of reliability of the representation, the term “generalized representation of the shape” can be defined as follows: the lowest level is representation based on the local disparity pattern of stimuli (i.e., not generalized); the subsequent is representation based on a more global shape information, irrespective of the orientation of surfaces; and the next higher level is representation based on the

global information of the shape, irrespective of both the orientation of surfaces and depth position. Similarly, the term “generalized representation of the orientation of 3D surface” can be defined as follows: the lowest level is representation based on the local disparity pattern of stimuli; the next level is representation based on more global information, irrespective of the shape; the subsequent level is representation based on global information that irrespective the shape and depth position. In our experiment, RDS was used to depict horizontally or vertically positioned convex and concave stereoscopic surfaces; the shape and orientation of these stimuli cannot be judged by monocular cues, but can be judged by binocular cues. The BOLD signal, as measured by functional magnetic resonance imaging (fMRI), can be used to indirectly reflect the underlying neural activity of cortices (Logothetis et al., 2001). We tested the global patterns of each area by classification using MVPA. The overall plan of the study was the following. First, the selectivity of neurons to shape and orientation of surface depicted by binocular disparity in each area was assessed. Second, by using stimuli of the same shape (convex or concave) or orientation (horizontal or vertical) defined by different disparities, it was compared whether the patterns invoked by stimuli with different disparities and the same shape/orientation are similar. Correspondingly, in the experiment, we initially sought to identify the areas involved in representing the shape and horizontal–vertical orientation of stereoscopic surfaces defined by disparity. Accordingly, we investigated the retinotopic visual cortices (V1, V2, V3d, V3v, V3A, and V7), the higher ventral cortex LOC, the higher dorsal area [human middle temporal complex (hMT+)], the kinetic occipital area (KO), and the IPS areas [the ventral intraparietal sulcus (VIPS), the parieto-occipital intraparietal sulcus (POIPS), and the dorsal intraparietal sulcus (DIPS)]. Because neurons selective to binocular disparity exist at the multiple levels of the visual hierarchy, it is possible that most regions of interest (ROIs) would be selective to the convex–concave shape and horizontal–vertical orientation. Same-type stimuli classification was used to test this. Furthermore, we wanted to test whether the selectivity relates to disparity information per se or more generalized processing independent of low-level retinotopic disparity information. The dorsal areas V3A and KO may be two of the candidates that be involved in a more generalized representation of the shape. This could be attributed to the belief that coarse stereopsis is processed in the dorsal stream and fine stereopsis is processed in the ventral stream (Neri et al., 2004; Uka & DeAngelis, 2006; Roe et al., 2007; Schiller et al., 1990). Coarse stereopsis can be used to guide the vergence movements of the eyes. During this process, it is possible that some abstract information regarding the shape is required. Traditionally, shape information is considered to be processed in ventral stream (Zeki, 1978; Mishkin et al., 1982). Therefore, the exchange of information between the ventral and dorsal streams and is required. Moreover, a more generalized representation of objects that can be shared between the dorsal and

ventral streams may be involved. Furthermore, 3D information has been shown to be exchanged between the two streams. Van Dromme et al. (2016) provided causal evidence for the flow of visual 3D information between the dorsal stream and the ventral streams. V3A and KO are considered to be parts of the anatomical pathway where the functionally different ventral and dorsal streams exchange information (Takemura et al., 2016). In addition, Dövencioğlu et al. (2013) demonstrated that KO plays an important role in integrating disparity and shading cues of 3D structure perception, suggesting a more generalized representation of depth structure in the dorsal stream. Therefore, V3A and KO are two of the candidate areas for generalized shape representation. For orientation representation, IPS areas are expected to be involved in the more generalized representations of horizontal–vertical orientation because previous studies have shown that the neurons in the posterior IPS areas in monkeys and humans are selective to orientation via monocular cues, such as texture gradients and perspective, as well as via binocular cues, such as binocular disparity (Naganuma et al., 2005; Rosenberg et al., 2013; Shikata et al., 2001; Tsutsui et al., 2001).

2.3 Methodology

2.3.1 Participants

A total of 11 participants were recruited for the fMRI experiments. Of these, three participants were excluded due to poor performance on judging convex and concave stereoscopic surfaces. “Acceptable performance” was defined as an accuracy of ≥75% for each stimulus type. Of the remaining eight participants, seven were male and one was female. One of the males was left-handed, whereas the other seven participants were right-handed. All participants had normal or corrected-to-normal vision. None had any history of mental illness or neurological disease. Their ages ranged from 22 to 33 (mean ± SD, 24.6 ± 3.7) years. Participants were remunerated for their participation.

This study was performed in accordance with the recommendations of Human Research Ethics Committee of the Kochi University of Technology, and written informed consent was obtained from all participants. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Human Research Ethics Committee of the Kochi University of Technology.

2.3.2 Stimuli

Stereoscopic stimuli were presented on a screen using a pair of JVC D-ILA video projectors. A linear polarized filter was placed in front of each projector. The images from the two projectors were superimposed into one image and projected onto a translucent screen inside the bore of the magnet. The participants wore polarized glasses and viewed stimuli through a slanted mirror (angled at 45°) above the head coil. The optical distance from the midpoints of the two eyes to the screen was 71 cm. The screen resolution was set at 1024 ×768 pixels and the refresh rate was 120 Hz.

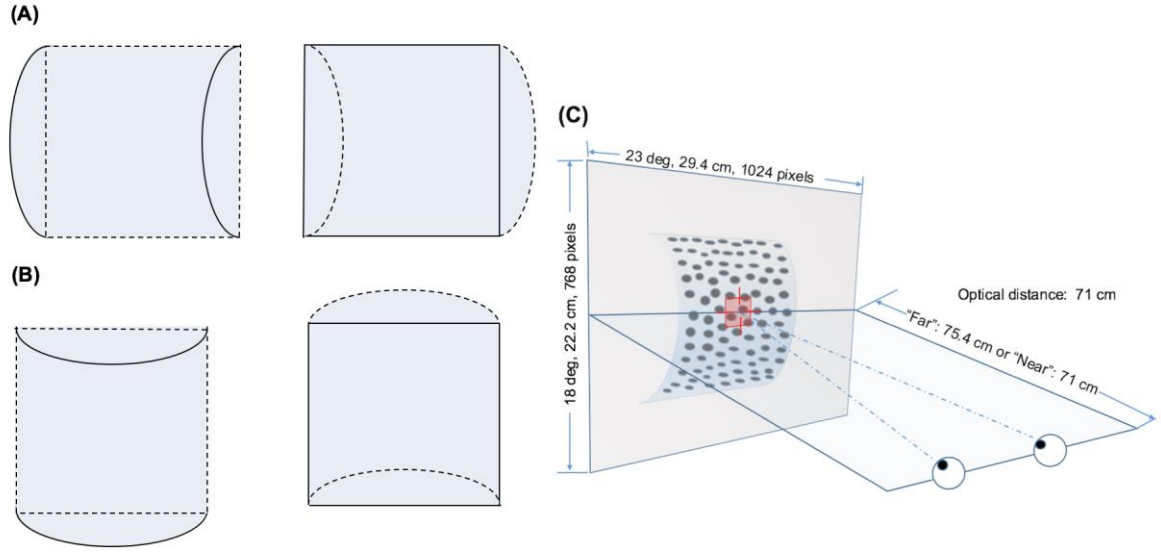

Figure 2.1 Schematic illustration of the stimuli. (A) Horizontal hemi-cylindrical convex (left) and concave (right) surfaces. (B) Vertical hemi-cylindrical convex (left) and concave (right) surfaces. (C) Experimental set up (the example is a convex near surface).

The RDS stimuli comprised random black and white dots generated using Psychotoolbox 3 in MATLAB (The MathWorks). They depicted four surface types defined by disparity as follows: the horizontal convex and concave curved surfaces (Figure 2.1A) and vertical convex and concave curved surfaces (Figure 2.1B). Each surface covered a square of 14° ×14° and was presented on a mid-gray rectangular background (23° ×18°). The density of the stereogram was 4 1 dots/deg^2; all the dots were of the same size and the diameter of the dots was 0.14°. The radius of the surface was 4.4 cm. The horizontal disparity of dots was calculated based on the depth position and there was no vertical disparity between corresponding dots. A fixation marker comprising a hollow square with a side length of 0.5° and horizontal and vertical nonius lines with a length of 0.375° was displayed at the center of the screen to help participants maintain eye vergence.

2.3.3 Experimental design

Four stimulus types, namely horizontal convex, horizontal concave, vertical convex, and vertical concave hemi-cylindrical surfaces, were presented during the experiment at two different depth positions (“near” or “far”) behind the fixation marker. The distance from the nearest part of the surface and fixation marker was simulated as 0 cm (“near” condition, uncrossed, absolute disparity of nearest part: 0°) or 4.4 cm (“far” condition, uncrossed, absolute disparity of nearest part: 0.30°). Stimuli selected from a set of eight conditions (4 types ×2 positions) were presented using a bl ock design (Figure 2.2). To avoid adaptation and maintain neuronal activation, the stimulus in each block was flashed on and off repeatedly for 0.5 s. Each time a stimulus pattern was shown, the random dots were regenerated. Each stimulus block, which comprised one of the eight conditions and lasted 15 s, was repeated twice in a random order per run. After each stimulus block, the participant was required to make a judgment of the shape of the surface by pressing the corresponding button on a keypad. The judgment was followed by a 6 s fixation block. There was no fixation block following the final judgment block. Each run began with a 12 s fixation block. The total time taken for one run is calculated as follows: (1) the time of all stimulus blocks is 15 ×8 ×2 = 240 s; (2) the time of all blocks for judgment is 3 ×8 ×2 = 48 s; and (3) the time of fix ation blocks is 12 + 6 × (8 × 2 − 1) = 102 s. Therefore, each run lasted a total of 240 + 48 + 102 = 390 s. Participants were required to observe the fixation marker and avoid any head movement during all runs. Any runs with excessive head movement were discarded. Excessive head movement was defined as head movement of >2 mm or head rotation of >2° during each run. Overall, participants were able to maintain their head static and at least six usable echo-planar imaging (EPI) scans were obtained for each participant.

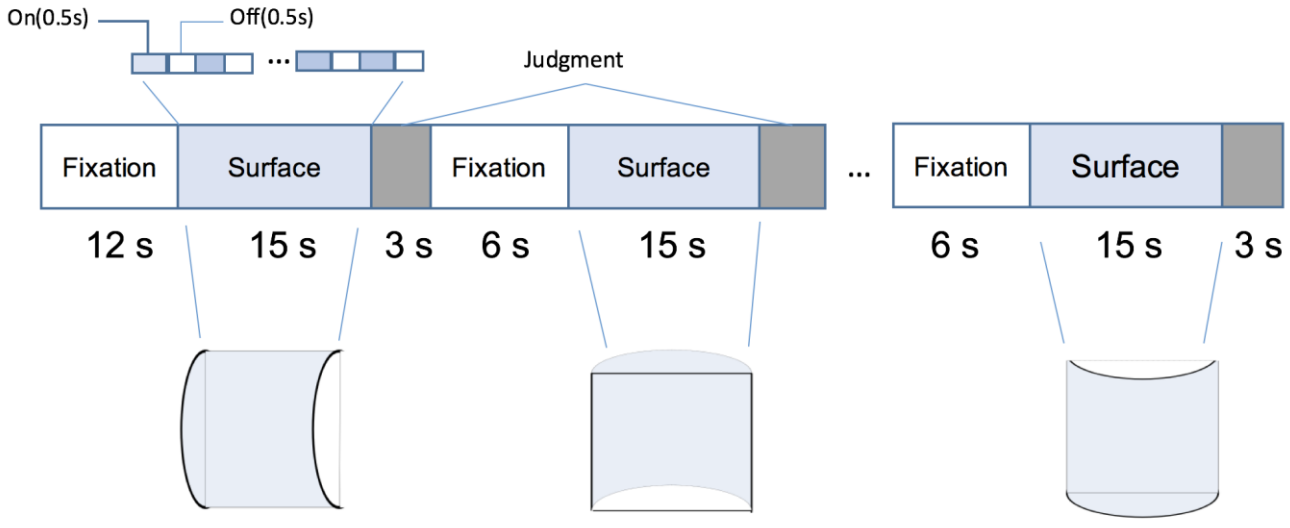

Figure 2.2 Illusion of the fMRI block design.

2.3.4 Data acquisition

Imaging was performed using a 3 Tesla Siemens Verio MRI scanner with a 24-channel multi-phase array head coil at the Brain Communication Research Centre of the Kochi University of Technology. Participants' heads were fixed with foam padding to reduce movement. For each participant, a high resolution T1-weighted anatomical scan (1 ×1 ×1 mm) was obtained to construct the exact inflated and flattened cortical surface. For the experimental scans, BOLD signals were measured using an EPI sequence [echo time (TE): 30 ms; repetition time (TR): 3000 ms; number of volumes per run: 130; slice thickness: 3 mm; slice acquisition order: interleaved] from 35 slices covering the visual cortex, posterior parietal cortex, and posterior temporal cortex. In addition, a T2-weighted structural image was obtained in a run of 2.5 min and it was recorded at the same position as the slices of the corresponding EPI data in one session. Structural data were used as reference slices for 3D motion correction of EPI data as well as co-registration between T1-weighted anatomical images and EPI images. Following co-registration between anatomical and functional data in native anatomical space, all data were converted to Talairach coordinates.

We measured the ROIs (Figure 2.3) for each participant using standard procedures in separate sessions prior to the main experiment. Retinotopically organized visual areas, namely V1, V2, V3d, V3v, and V3A, were localized by rotating wedge stimuli and expanding concentric rings (Sereno et al., 1995; DeYoe et al., 1996; Warnking et al., 2002). Area V7 was defined as the region anterior and dorsal to V3A with a lower visual field quadrant representation (Tootell et al., 1998; Tyler et al., 2005). Moreover, we identified some higher ventral areas including the LOC, the higher dorsal areas hMT+ and KO, and areas along the IPS including VIPS, POIPS, and DIPS using special independent localizers. LOC was identified as a set of continuous voxels in the lateral occipito-temporal cortex that showed significantly stronger activation ($p < 10^{-4}$) to intact versus scrambled images of objects (Kourtzi et al., 2000; Kourtzi et al., 2001). hMT+ was defined as a set of continuous voxels in the lateral temporal cortex that demonstrated significantly stronger activation ($p < 10^{-4}$) to a set of coherent outward and inward moving dots compared to static dots (Zeki et al., 1991). KO was identified as a set of voxels that responded significantly stronger activation ($p < 10^{-4}$) to motion-defined contours than the transparent motion of a field of black and white dots (Dupont et al., 1997; Zeki et al., 2003). Stimuli comprising nine randomly connected lines were used to locate the IPS areas (VIPS, POIPS, DIPS). These areas were identified by contrasting activity to 3D shapes, which was produced by rotating stimuli in depth, versus activity to 2D shapes, which was produced by

moving stimuli along a frontoparallel plane (Vanduffel et al., 2002).

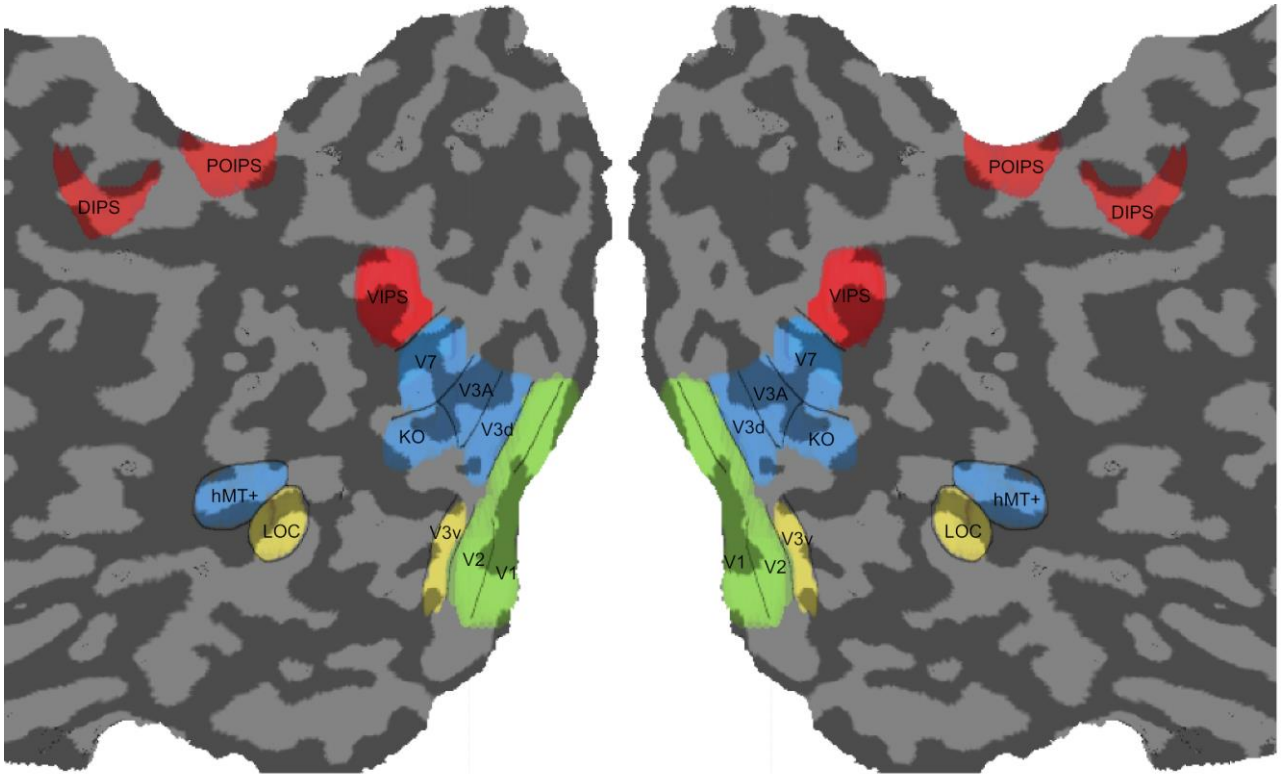

Figure 2.3 Illustration of a flattened cortical surface with ROIs superimposed. The retinotopic areas (V1, V2, V3d, V3v, V3A), higher ventral area (LOC), higher dorsal areas (KO, V7, hMT+), and intraparietal sulcus areas (DIPS, POIPS, VIPS). Sulci are indicated in dark gray whereas gyri are indicated in light gray. ROIs were defined by standard localizers in separate sessions (see the data acquisition section 2.3.4).

2.3.5 Data analysis

2.3.5.1 Pre-processing

Data processing and analysis were performed using the Freesurfer software package (Fischl, 2012), BrainVoyager QX (Version 2.8.4.2645, 64-bit; BrainInnovation, Maastricht, The Netherlands), MATLAB R2014a (The Mathworks), and SPSS Statistics 23 (IBM Inc.). Freesurfer was used to remove the scalp of the T1-weighted 3D anatomical image per participant and segment the remaining parts into different regions, such as those of the outer skull and inner skull. The white matter (WM) was separated from the other components and later used as a mask to segment WM and gray matter (GM) in BrainVoyager QX. The brain was extracted from the skull and other tissues. Thereafter, we used BrainVoyager QX to transform the extracted brain into standard Talairach space. The flattened cortical surface of each participant was generated by segmenting the cortical surface along the

GM/WM boundary, inflating the segmented GM, cutting along the calcarine sulcus, and flattening. The flattened cortical surface was used for visualizing functional maps and delineating ROIs, which were later used in the MVPA procedure. For the EPI data, 3D motion correction was performed (Woods et al., 1998) referencing the T2-weighted image obtained at the beginning of each session. No spatial smoothing was performed. Finally, co-registration was performed to align the functional EPI data to the T1-weighted anatomical data, and the functional EPI data were transformed into Talairach space.

2.3.5.2 ROI-based univariate analysis

Univariate analysis was performed to assess whether the overall activity of the neurons in each selected ROI was higher for all stimuli compared with that at the fixation baseline. We calculated the "percent signal change" for the stimuli versus the fixation baseline. The fMRI time courses were shifted by two volumes (6 s) to account for the hemodynamic delay of the BOLD signal. The percent signal change values were averaged across all runs and all participants.

SPSS Statistics 23 was used for statistical hypothesis testing. One-sample t-tests were used to assess whether percent signal change was significantly higher than chance level for each ROI.

2.3.5.3 ROI-based MVPA

MVPA method is widely used to identify the global patterns of the cortical areas of the human brain and is useful for discovering function of the cortical areas. It allows the detection of subtle differences between the conditions of interest but it may bear unexpected pitfalls (Alizadeh, S. et al., 2017). We carefully performed a series of MVPA for the EPI data for each ROI in MATLAB. SVM was used as the classifier for MVPA. The implementation of SVM was provided by MATLAB. Typically, two types of classifications were performed. The first was a same-type stimuli classification wherein we trained and tested SVM on the same type of data. This was performed to investigate whether specific patterns were contained in the dataset. The second was a transfer classification wherein we trained and tested SVM on different types of data. This was performed to assess whether a pattern contained in one type of data was also contained in another type of data. This is a useful method for determining whether a specific pattern is shared by different datasets.

Our MVPA procedure is as follows. Due to the hemodynamic delay of the BOLD signal, the fMRI time courses were shifted by two volumes (6 s). For each ROI in Talairach space, voxels were selected from both hemispheres and sorted from the largest to smallest according to their response (t-statistic) to the stimulus conditions compared with the fixation baseline. Using the t-statistics from the "stimulus versus fixation baseline" contrast, the top 250 voxels were selected for all ROIs and all

participants. For ROIs of participants for whom 250 voxels were unavailable, the highest number of voxels available was used for the classification.

To estimate the value for each voxel of the ROIs used as input for SVM for each stimulus block, the average value of the first three volumes after stimulus onset (avg1) and the average value of the last two volumes before stimulus onset (avg2) corresponding to the fixation baseline (after shifting by two volumes) were calculated. The avg1–avg2 difference was then calculated. Thereafter, the time series differences were transformed into z-scores and transmitted to SVM for training and testing.

The leave-one-run-out cross-classification method was used to evaluate the performance of the MVPA classification, i.e., data were partitioned according to the run they belong to, and data from one run were used for testing, whereas data from other runs were used for training. This procedure was repeated with different partitions of the data. The classification accuracy for each ROI was calculated as the average classification accuracy of all cross-classifications. The classification accuracies of all participants were averaged for each ROI.

To estimate the baseline for statistical significance, we performed classification analysis with randomly permutated fMRI patterns for all ROIs, i.e., we randomized the correspondences between fMRI data and labels and performed classification similar to that performed for normal non-permutated data. This procedure was repeated 1000 times to create a distribution of classification accuracies. We used the 99.6th percentile (one-tailed, 12 ROIs) as the baseline for statistical significance.

2.4 Result

2.4.1 ROI-based univariate analysis

The average percent signal changes across all runs and all participants per ROI are shown in Figure 2.4. All ROIs showed a percent signal change >0 for the condition stimuli versus fixation baseline, indicating that the neurons from these ROIs were more active during the stimuli condition than in the fixation baseline condition. However, the percent signal change was not significantly greater than 0 for the dorsal area hMT+ and parietal areas POIPS and DIPS. The reason for the absence of a significant change by these ROIs could be the trend in which signal changes increasingly became weaker from the early visual areas to higher visual areas, that the signal changes in these three areas were relatively weak, and that our criterion for significance was strict (*$p \leqslant 0.004$). Regardless, they still showed an average signal change greater than 0.

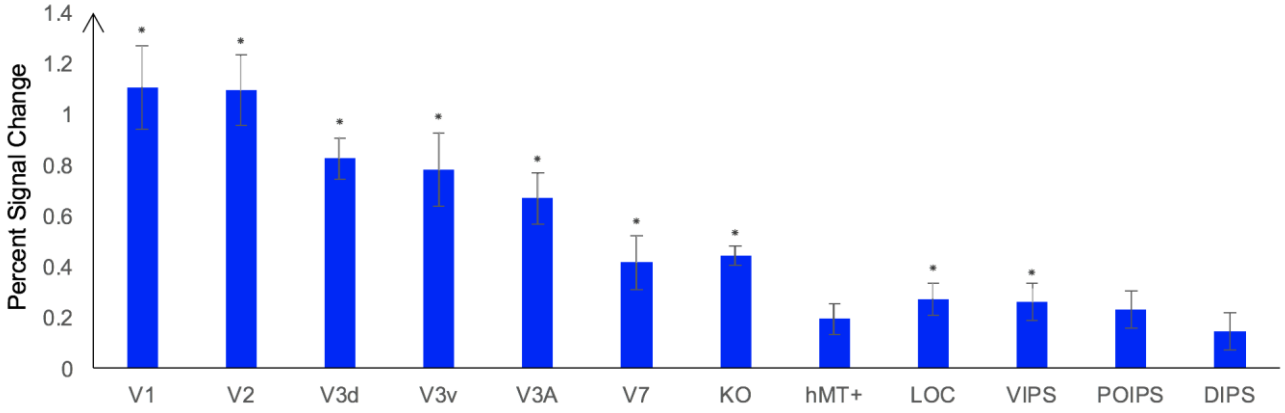

Figure 2.4 Percent signal change of each ROI. Error bars represent standard error of the mean across participants (n = 8). Asterisks indicate that the percent signal change is significantly greater than 0, as assessed by a t-test of group data (*p ⩽ 0.004).

2.4.2 ROI-based MVPA

2.4.2.1 Classification accuracy for the shape discrimination

We performed three types of the shape classifications (Figure 2.5) as follows: (i) same-type stimuli shape classification (denoted by green arrows); (ii) transfer shape classification of surfaces at the same depth position (denoted by red arrows); and (iii) transfer shape classification of surfaces at different depth positions (denoted by blue arrows).

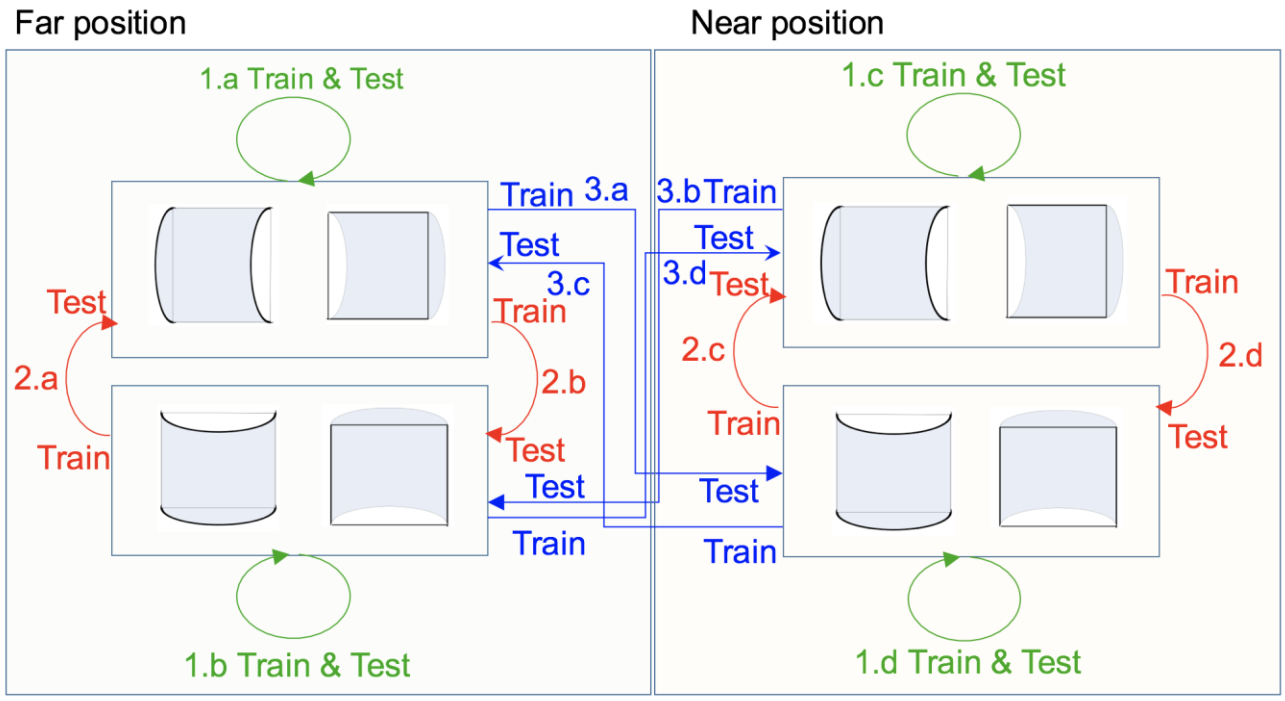

Figure 2.5 Illustration of the types of the shape classification performed. 1.a, 1.b, 1.c, and 1.d represent same-type stimuli shape classification. 2.a, 2.b, 2.c, and 2.d represent transfer shape classification of surfaces at the same depth position. 3.a, 3.b, 3.c, and 3.d represent transfer shape classification of surfaces at different depth positions.

(1) Same-type stimuli shape classification

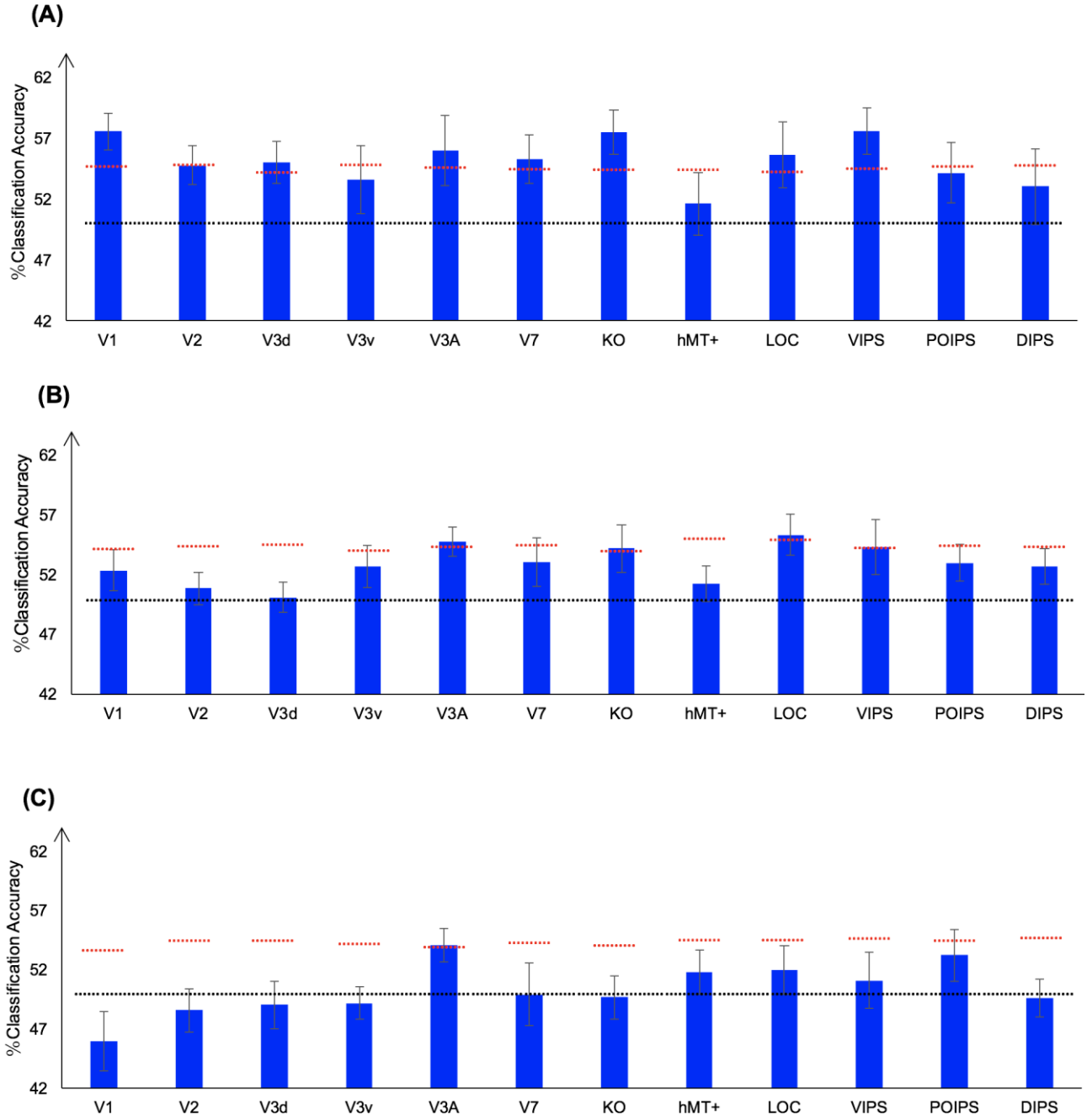

Figure 2.6 The accuracy of classification for convex versus concave discrimination. (A) The accuracy of classification for the shape discrimination for same-type stimuli. (B) The transfer shape classification of surfaces at the same depth position. (C) The transfer shape classification of surfaces at different depth positions. The red horizontal dotted lines indicate the baselines of statistical significance for ROIs. The locations of these lines indicate the upper 99.6% percentile of the distribution of the accuracy of classification of the permutated data. The black horizontal line indicates the chance level for two-class classification (50%). The error bars depict the standard error of the mean across subjects (n = 8).

To investigate whether ROIs are selective to the shape of convex and concave surfaces, “same-type

stimuli shape classification" was performed. Each stimulus exhibited the following three attributes: position (near or far), orientation (horizontal or vertical), and shape (convex or concave). Here, the "same-type stimuli shape classification" refers to trained and tested SVM using the same type of data, i.e., both the training and the testing data were of the same depth position (near or far) and orientation (horizontal or vertical). For example, SVM was trained on (horizontal, near) surfaces data and then tested using the same type of data (horizontal, near) surfaces. Similarly, we trained SVM on [(horizontal, far), (vertical, near), and (vertical, far)] surfaces separately, and then tested SVM on the corresponding types of data (see 1.a, 1.b, 1.c, 1.d of Figure 2.5). The average classification accuracy values across the four types of data of the eight participants are shown in Figure 2.6A.

All areas showed an average classification accuracy higher than chance level for the two-class classification (50%). The V1; V2; dorsal areas V3d, V3A, V7, and KO; ventral area LOC; and parietal area VIPS exhibited a classification accuracy higher than baseline of statistical significance, suggesting that the activity patterns of neurons in these areas are selective to the shape of curved surfaces.

(2) Transfer shape classification of surfaces at the same depth position

The ability of SVM to classify the shape with an accuracy higher than baseline of statistical significance is attributable to two main reasons. One is that the activity patterns of the neurons reflect the local disparity of surfaces. The local disparity between surfaces with a different shape was different; it is possible that the activity patterns of neurons in ROIs reflect these differences and that SVM used the different activity patterns for classification. The second possibility is that the activity patterns of neurons reflect a more generalized shape representation and SVM classified surfaces based on these patterns.

To assess whether each ROI is related to binocular disparity or more generalized processing of shape representation which is not related to the orientation of surface, the transfer shape classification of surfaces at the same depth position was performed.

There are three points that should be mentioned as follows: (1) the shape classification was performed by classifying whether a convex or concave surface was shown during the stimulus blocks; (2) the training and testing data were selected from blocks when stimuli were shown at the same depth condition; and (3) 'transfer' indicates that the data used for training and testing had different attributes in terms of orientation (horizontal or vertical). For example, we trained SVM using data in which horizontal surfaces were shown and then tested SVM using data in which

vertical surfaces were shown at the same depth position and vice versa. The classification is illustrated in 2.a, 2.b, 2.c, and 2.d of Figure 2.5. The average accuracy values for the four classification types across the eight participants are shown in Figure 2.6B. In the higher dorsal areas V3A and KO, ventral area LOC, and parietal area VIPS, the fMRI responses evoked by one type of surfaces (horizontal or vertical) could allow the shape classification of responses evoked by the other type of surfaces (vertical or horizontal, respectively). Because the disparity patterns of the surfaces used for training and testing were different, this result was probably related to more generalized processing of the shape.

To further investigate whether generalized representation is invariant of depth position, we performed the transfer shape classification of surfaces at different depth positions.

(3) transfer shape classification at of surfaces different depth positions

In this type of classification, SVM was trained to classify the shape of surfaces. “Transfer” indicates that the data used for training and testing have different attributes in terms of both depth position and orientation. For example, we trained SVM using data with horizontal surfaces in the far position and tested SVM using data with vertical surfaces in the near position. These classifications are illustrated by arrows 3.a, 3.b, 3.c, and 3.d of Figure 2.5. The average accuracies for the four types of classification across eight participants are shown in Figure 2.6C.

As shown, only the classification accuracy for V3A was higher than baseline of statistical significance (classification accuracy: 54.05%; baseline of statistical significance: 53.92%). This indicates that V3A is related to a more generalized representation of shape and is invariant to depth-position.

Alternatively, the near–far information between the center and surrounding of the surfaces could support the all the three types of shape classification. Moreover, for “the same-type stimuli shape classification” and “transfer shape classification of surfaces at the same depth position”, the different depth positions of peaks of surfaces with a different shape could be used for classification. To exclude this possibility, we investigated the classification accuracy for near versus far surface discrimination. Figure 2.7 illustrates the near–far classification performed. We performed same-type stimuli near–far classification. SVM was trained to classify a near versus far position of the surfaces. “Same-type stimuli” indicates that the data used for training and testing exhibited the same shape (convex or concave) and same orientation (horizontal or vertical). For example, we trained and tested SVM using data with horizontal surfaces with minus curvature. The average classification accuracies

for the four types of data across all participants are shown in Figure 2.8.

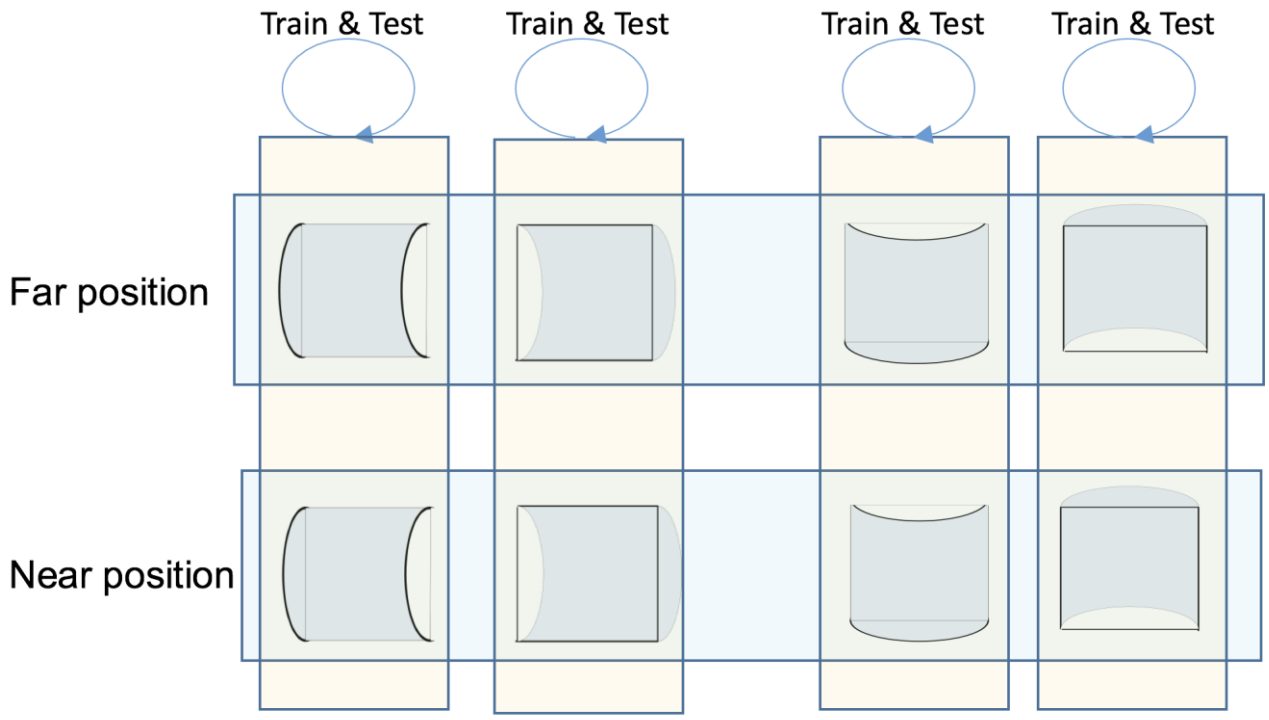

Figure 2.7 Illustration of the near–far classification. We trained and tested SVM using the same type of data.

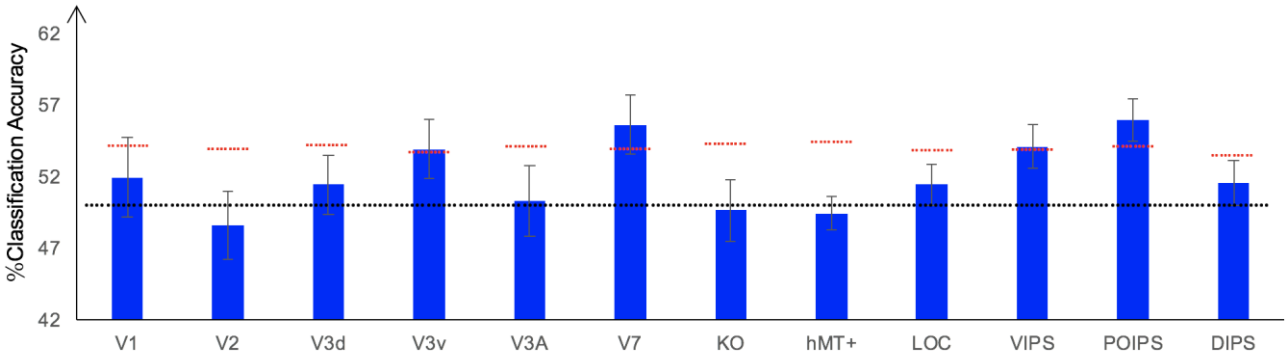

Figure 2.8 The accuracy of classification for near versus far discrimination using same-type stimuli. The red horizontal dotted lines indicate the baselines of statistical significance for ROIs. The locations of these lines indicate the upper 99.6% percentile of the distribution of the accuracy of classification of the permutated data. The black horizontal line indicates the chance level for two-class classification.

As shown, the average classification accuracy for V3A was approximately that of chance level for two-class classification, indicating that near–far surfaces cannot be classified by V3A, at least in our experimental setting. By combining our findings thus far, we can conclude that the higher-than-baseline accuracy for the shape classification in V3A was not likely to be based on the activity

patterns of neurons that reflect near–far information between the center and side of the stimulus (because we cannot classify near–far surfaces by V3A in our experimental setting); rather, it is likely related to activity patterns of neurons that reflect the generalized shape representation. This result is consistent with our hypothesis that the dorsal area V3A is a candidate involved in generalized shape representation.

The results showed high accuracy in the three types of the shape discrimination in V3A. The information used for the discrimination, however, may be different: (1) for the "same-type stimuli shape classification," it may to be explained by local disparities (or zero-order disparity) because of the training data and testing data were of identical type. (2) for the "transfer shape classification of surfaces at the same depth position," it is may be classified according to the different depth positions of the peaks of surfaces with the different shape or near–far information between the center and surroundings of the surfaces. However, we also checked "same-type stimuli near-far classification" for which disparities were different between near and far surfaces. The result showed that in our experimental condition, the accuracy of the classification was approximately that of chance level in V3A. Considering this finding, we can infer that it was unlikely be classified by these two reasons; and (3) for the "transfer shape classification of surfaces at different depth position", the disparity of surfaces for training (e.g., horizontal surfaces in the near depth position) and disparity of surfaces for testing (e.g., vertical surfaces in the far depth position) were quite different. Hence, it is unlikely to be classified by local disparities. In addition, V3A is on the dorsal visual stream, and this visual stream is related to coarse stereopsis. It is likely that V3A was involved in the representation of the shape in a more abstract (coarse) manner, which did not depend on the orientation of the surface and depth position. Therefore, V3A showed high accuracy in this classification. Notably, the high accuracy demonstrated by V3A showed in our classification does not indicate that the ventral stream areas were not involved in the generalized representation of shape. There is an exchange of information between dorsal and ventral streams. We observed that V3A showed statistical significance in all the three types of shape classification. In conclusion, V3A is a crucial part of the process related to the more generalized shape representation and the results are consistent with those reported in previous studies.

2.4.2.1 Classification accuracy for the shape discrimination

In addition to the representation of the shape defined by disparity, we investigated the representation of orientation of stereoscopic surfaces among ROIs. Accordingly, ROI-based MVPA was performed.

The following three types of horizontal versus vertical surface classifications were conducted (Figure 2.9): (i) the same-type stimuli horizontal–vertical surfaces classification; (ii) the transfer horizontal–vertical classification of surfaces with different shapes at the same depth position; and (iii) the transfer horizontal–vertical classification of surfaces with different shapes at different depth positions.

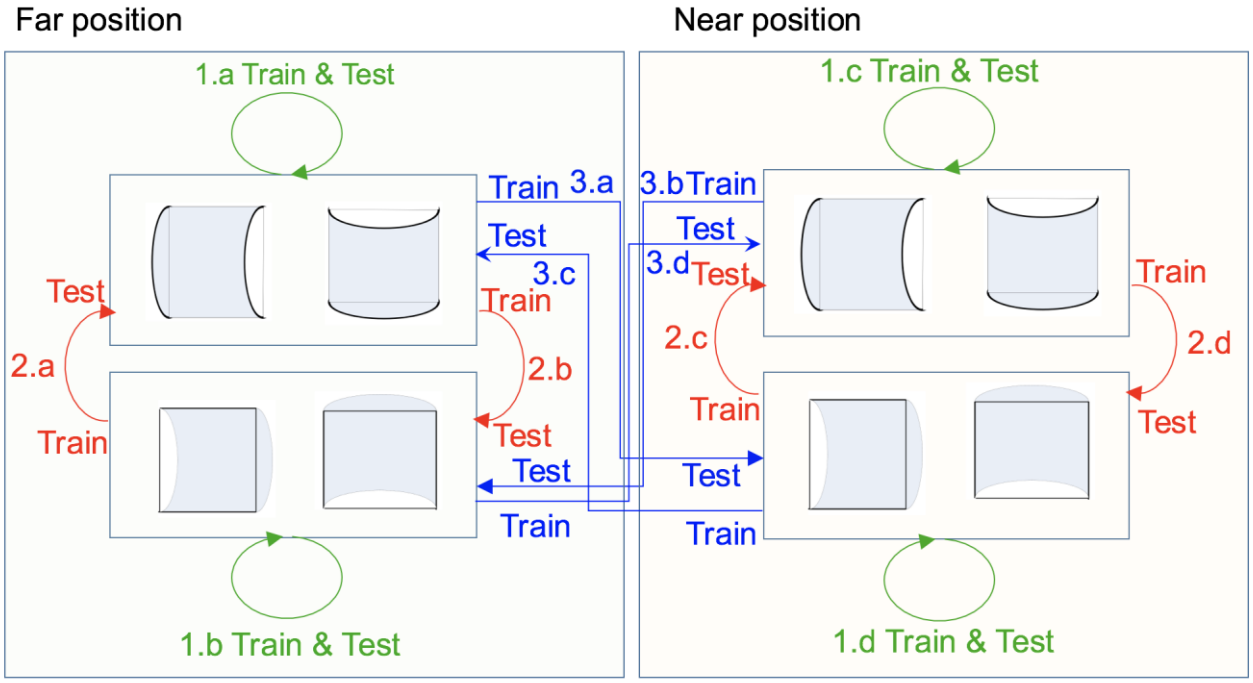

Figure 2.9 Illustration of the horizontal–vertical classification. 1.a, 1.b, 1.c, and 1.d indicate same-type stimuli horizontal–vertical surfaces classification. 2.a, 2.b, 2.c, and 2.d indicate transfer horizontal–vertical classification of surfaces with different shapes at the same depth position. 3.a, 3.b, 3.c, and 3.d indicate transfer horizontal–vertical classification of surfaces with different shapes at different depth positions.

(1) Same-type stimuli horizontal–vertical surfaces classification

We trained SVM to classify the orientation of surfaces, i.e., to judge whether surfaces were horizontal or vertical. “Same-type stimuli” indicate that both the training and the testing data were of the same shape (convex or concave) and depth position (near or far). For example, we trained and tested SVM using data with horizontal–vertical surfaces with convex shape in the near position. This classification is illustrated in 1.a, 1.b, 1.c, and 1.d of Figure 2.9. The average classification accuracies for all data types across the eight participants are shown in Figure 2.10A.

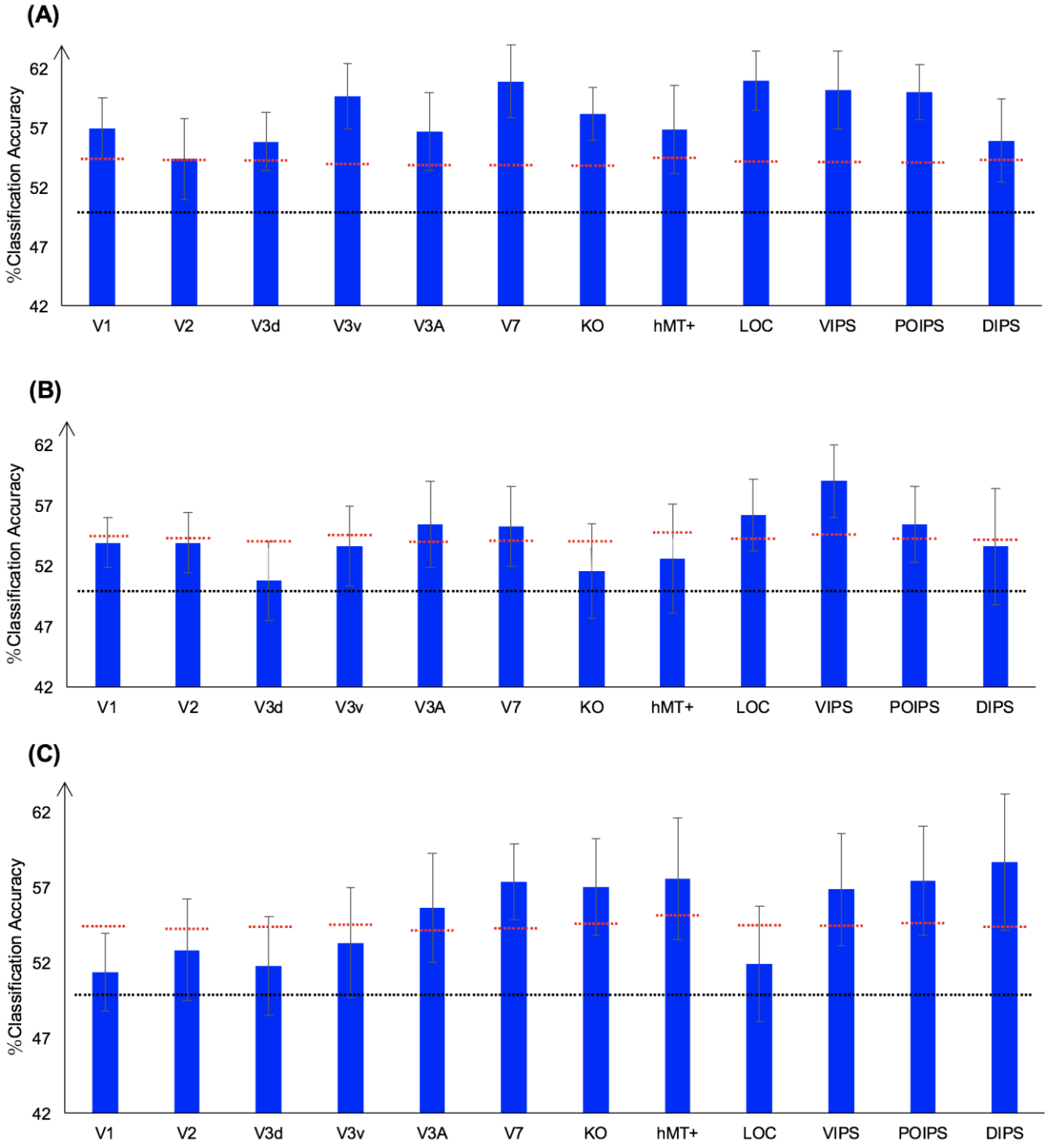

Figure 2.10 The accuracy of classification for horizontal versus vertical discrimination. (A) The accuracy of classification for horizontal versus vertical discrimination using same-type stimuli. (B) The accuracy of classification for transfer horizontal–vertical classification of different shape surfaces at the same depth position. (C) The accuracy of classification for transfer horizontal–vertical classification of different shape surfaces at different depth positions. The red horizontal dotted lines indicate the baselines of statistical significance for ROIs. The locations of these lines indicate the upper 99.6% percentile of the distribution of the accuracy of classification of the permutated data. The black horizontal line indicates the chance level for two-class classification.

The accuracies of classification for distinguishing horizontal versus vertical orientation among all areas were higher than baseline of statistical significance, indicating that the activity patterns of the neurons in all ROIs contain robust orientation information.

(2) Transfer horizontal–vertical classification of surfaces with different shapes

This classification included two types as follows: the "transfer horizontal–vertical classification of surfaces with different shapes at the same depth position" and the "transfer horizontal–vertical classification of surfaces with different shapes at different depth positions."

The reasons for SVM showing a horizontal–vertical classification accuracy higher than baseline of statistical significance for same-type stimuli horizontal–vertical surfaces may be the following. First, the disparity patterns of horizontal surfaces and vertical surfaces were different. The activity patterns of neurons could reflect this disparity, and SVM used of the different patterns of the neurons for the classification. Second, the activity patterns of neurons could reflect horizontal and vertical orientation representation processing, i.e., these activity patterns reflect the representation of horizontal–vertical orientations of surfaces and SVM classified the surfaces based on these patterns. To investigate whether the high accuracy for each ROI depended on disparity or more generalized horizontal–vertical orientation representation irrespective of shape, "transfer horizontal–vertical classification of surfaces with different shape at the same depth position" was performed. SVM was trained to classify whether a stimulus shown to a participant was horizontal or vertical. "Transfer" indicates that the stimuli used for training and testing exhibited different shapes but the same depth position. For example, data of surfaces with a convex shape shown in the near position were used to train SVM and data of surfaces with a concave shape shown in the near position were used to test SVM. The classification is illustrated in 2.a, 2.b, 2.c, and 2.d of Figure 2.9. The average accuracies of classification for the four types of classification across all participants are shown in Figure 2.10B. In the dorsal areas V3A and V7, higher ventral area LOC, and parietal areas VIPS and POIPS, the fMRI responses evoked by one type (minus or plus) of horizontal–vertical surface can allow the classification of the response evoked by another type (plus or minus, respectively) of horizontal–vertical surface at the same depth position. This finding suggests that these areas are related to the generalized processing of orientation of stereoscopic surfaces irrespective of disparity types.

For the "transfer horizontal–vertical classification of surfaces with different shapes at different depth positions," SVM was trained to classify the horizontal–vertical orientation of the surface of a stimulus shown to participants. "Transfer" indicates means that the data used for training and testing had different attributes in shape (convex or concave) and depth (near or far). For example, SVM was trained using data with a surface with a convex shape in the near position and SVM was tested using data with a surface with a concave shape in the far position. These classifications are shown by

arrows 3.a, 3.b, 3.c, and 3.d in Figure 2.9. However, there is other information that could support this type of horizontal–vertical classification, namely there are some common disparity patterns between the surfaces of different shapes and the same orientation at different depth positions. For example, horizontal concave surfaces in the near position and horizontal convex surfaces in the far position share a similar pattern at the horizontal center line where the two surfaces may converge if shown simultaneously. Similarly, vertical concave surfaces in the near position and vertical convex surfaces in the far position share a similar pattern at the vertical center line where the two surfaces may touch if they are shown at the same time. Figure 2.11 shows a schematic illustration of these examples. SVM could use these patterns to classify the horizontal–vertical orientation of surfaces. Therefore, whether the representation is invariant of depth position cannot be verified in this case. The average accuracies of all classifications across the eight participants are shown in Figure 2.10C. As shown, in addition to the areas (except the LOC) with an accuracy of classification higher than baseline of statistical significance for "transfer horizontal–vertical classification of with different shape at the same depth position," those of the dorsal areas KO and hMT+ and the parietal area DIPS are higher than baseline of statistical significance. For the LOC, accuracy is higher than the baseline of statistical significance in "transfer horizontal–vertical classification of surfaces with different shapes at the same depth position" but not for "transfer horizontal–vertical classification of surfaces with different shapes at different depth positions." This is possible because the LOC is selective to depth position. Preston et al. (2008) showed that the LO, a sub-region of the LOC, contains information regarding the depth position of planes depicted by an RDS.

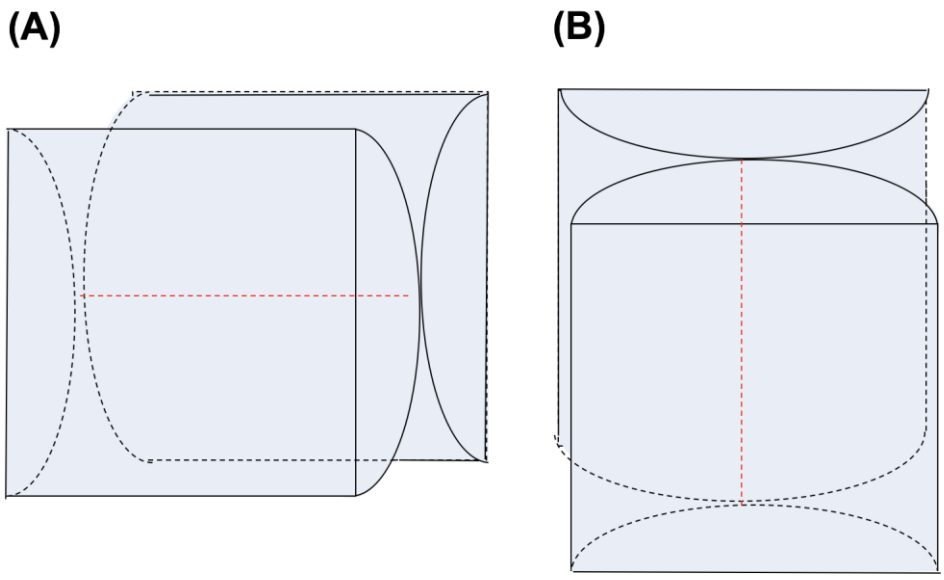

Figure 2.11 A schematic illustration of two surfaces of different shapes of the same orientation at different depth positions that share some similar disparity patterns. (A) A horizontal concave surface in the near position and a horizontal convex surface in the far position are shown. These two surfaces share a similar disparity pattern around the position shown by the horizontal red dotted line.

(B) Vertical concave surface in the near position and vertical convex surface in the far position. These two surfaces share a similar disparity pattern around the vertical dotted red line.

In summary, although we were unable to confirm that the orientation representation is invariant of depth position, we can conclude that the dorsal areas V3A and V7, the higher ventral area LOC, and the parietal areas VIPS and POIPS are somewhat related to the generalized representation of horizontal–vertical orientation of surfaces. In addition, because in the LOC, accuracy in "transfer horizontal–vertical classification of surfaces with different shapes at different depth positions" was lower than the baseline for statistical significance, we can confirm that the orientation representation in the LOC is not invariant with depth position.

2.5 Discussion

In the present study, we used the MVPA method and fMRI data to investigate the shape and orientation representations of 3D surfaces defined by binocular disparity. By comparing a series of the accuracies of classification across ROIs, we found that all types of shape classification—including the "same-type stimuli shape classification," "transfer shape classification of surfaces at same depth position," and "transfer shape classification of surfaces at different depth positions"—showed an accuracy of classification higher than baseline of statistical significance in the dorsal area V3A. For horizontal–vertical orientation classification, the accuracies of classification for both "same-type stimuli horizontal–vertical classification" and "transfer horizontal–vertical classification of surfaces with different shapes at same depth position" were significant in the dorsal areas V3A and V7, higher ventral area LOC, and parietal areas VIPS and POIPS. In summary, these results indicate that the dorsal area V3A is related to the generalized representation of the shape and that some dorsal and ventral areas as well as parts of IPS are related to the generalized horizontal–vertical orientation representation of surfaces defined by disparity.

2.5.1 Comparison with earlier studies of shape representation

In the "two-stream" theory of cortical areas, visual information is processed progressively from the early visual areas to higher areas and divided into two anatomically and functionally separate streams after V1. Each stream processes visual information in a hierarchical manner, with each area processing information based on the results of previous areas (Zeki, 1978). Mishkin and Ungerleider (1982) first conceptualized this theory based on lesion research of non-human primates. The ventral stream has been termed the "what" stream whereas the dorsal stream has been termed the "where"

stream, as the ventral stream is related to an object's shape and identity whereas the dorsal stream is related to object's location and spatial relationships. According to this perspective, 3D shape representation should exist on the ventral stream. More recently, a revised "two-stream" theory was offered by Goodale et al. (1992). Rather than viewing both streams as contributing to conscious visual awareness, they argue that only the ventral stream contributes to conscious vision (known as the "perception" stream). Information in the dorsal pathway is used for the unconscious control of action, such as the movement of the body guided by visual input (known as the "action" stream). This functional separation of the ventral and dorsal streams is supported by various research including single-unit recording (Sakata et al., 1995; Fogassi et al., 2001), neuropsychological studies (Goodale et al., 1991; James et al., 2003), behavioral psychophysics (Ganel et al., 2003), and functional imaging (Culham et al., 2003; Monaco et al., 2014), and among others. From this view, shape representation can exist on both the ventral and dorsal pathways with representation in the ventral stream serving as perception and in the dorsal stream as visually guided action. There is ample evidence that 3D shape information from monocular cues is processed both in the ventral and dorsal visual streams in humans (Orban et al., 1999; Paradis et al., 2000; Taira et al., 2001; Georgieva et al., 2008). In macaques, stereo information has been found in both the dorsal and ventral visual stream (Parker, 2007). Further, in an fMRI study of patients suffering from visual object agnosia due to a ventral cortex lesion (dorsal cortex is intact), Freud et al. (2017) found that the dorsal cortex can mediate object representations that are dissociable from object representations in the ventral stream. Together with other evidence, they claimed that representations in the dorsal stream mediate the processing of object-related structural information, but are insufficient for normal object perception. There is growing evidence that independent object-selective representations exist in the dorsal visual pathway, which was once regarded as a function existing solely in the ventral visual pathway (Freud et al., 2016). However, there is no complete agreement on the distinction between the two pathways and the role played by each area on the pathways. Ventral "perception" and dorsal "action" separation has been challenged by findings that dorsal areas are also involved in shape representation for tasks that do not involve visually guided action (Konen and Kastner, 2008; Lehky and Sereno, 2007; Sereno et al., 2002).

For studies of single areas of the visual cortex, one challenge is to reveal what information is encoded in that area and to determine the manner in which its outputs responding to local simple elements are progressively transformed to encode the critical features of spatially extensive objects, such as shape and orientation, in other areas. Binocular disparity is a strong cue for 3D shape perception. Disparity-selective neurons are widely distributed throughout the visual cortical areas,

from as early as the striate cortex (V1 or primary visual cortex) to extra-striate and higher areas like the posterior parietal and inferior temporal cortices in both the dorsal stream and ventral visual streams (Neri et al., 2004). In the early visual area V1, neurons show selectivity for absolute disparity of RDS stimuli (Cumming & Parker, 1997). Neurons in V2 show selectivity for absolute and relative disparities (Thomas et al., 2002) and can transform absolute disparity to relative disparity during 3D vision, which is important for invariant object recognition (Grossberg et al., 2011). Besides the V3A, the dorsal regions including V3B/KO and V7 are involved in processing disparity-defined depth (Preston et al., 2009). For functional connections, dorsal pathway activation is correlated with the anterior ventral pathway activation (Sim et al., 2015). The neurons' selectivity for binocular disparity creates a foundation for ROIs showing accuracy for classifying stereoscopic shapes or orientations above baseline of statistical significance in our experiment, whereas the wide distribution of these neurons suggests that several areas, from the early visual areas to higher areas, offer a high-level accuracy of classification.

In the present study, we first examined the same-type stimuli shape classification. The neurons selective to the shape were found in both the dorsal (V1, V2, V3d, V3A, V7, KO, VIPOS) and ventral (V1, V2, LOC) streams. This finding is consistent with the two-stream theory, in which shape representation in the ventral stream mediates shape perception while shape representation in the dorsal stream mediates visually guided actions. However, we cannot be certain whether the shape representations in the dorsal areas are involved in visually guided actions or not, although our results support that this representation can exist in the dorsal areas. Therefore, this question thus requires further research.

Besides investigating the selectivity to the shape of surfaces defined by disparity, we investigated whether the shape representation was caused by a low-level binocular disparity or more generalized representation, which may relate to perception or visually guided actions. This was performed by the "transfer shape classification of surfaces at the same depth position" and "transfer shape classification of surfaces at different depth positions" analyses; the analyses revealed that all classification accuracies in the dorsal area V3A were higher than baseline of statistical significance, indicating that V3A is involved in the more generalized representation of shape curvature that does not depend on local disparity pattern (orientation and position in depth).

To the best of our knowledge, there is no direct research on the generalized representation of the shape of surfaces solely defined by disparity. However, there is some related research. Dövencioğlu et al. (2013) adopted convex "bumps" and concave "dimples" defined by shading, binocular disparity, and their combination. By comparing pattern classifications in different visual cortical

areas, they found that fMRI BOLD signals in the dorsal visual area V3B/KO were more discriminable when disparity and shading concurrently signaled depth, indicating the integration of these two cues. In a cross-cue transfer test, they found that the fMRI BOLD signal evoked by one type of cue could support the classification of a signal evoked by another type of cue, which indicates the generalized representation of shapes in the dorsal visual cortex by combining qualitatively different information with 3D perception. However, the results of the present study differ from those of Dövencioğlu et al. There are several possible reasons for this difference. First, we focused on 3D shape from binocular disparity whereas Dövencioğlu and colleagues focused on 3D shape from binocular disparity and shading. Our research revealed a common activity pattern of the neurons caused by the surface of same shape from different disparity patterns, whereas Dövencioğlu et al.'s research revealed a common activity pattern of the neurons caused by the same shape from different cues. Second, the convex–concave shapes used in our experiment and those in Dövencioğlu and colleagues' experiment were different. We adopted horizontal-positioned and vertical-positioned hemi-cylindrical convex–concave surfaces, while they used convex "bumps" and concave "dimples" of hemispheric shapes. Nevertheless, our finding that V3A is involved in a more generalized shape representation among ROIs does not contradict that of previous research. In macaques, clusters of disparity-selective neurons have been found in V3A (Anzai et al., 2011; Hubel et al. 2015). Using a 7T fMRI scanner, Goncalves et al. (2015) showed that disparity-selective neurons in the visual cortical areas were clustered and that this organization persisted across imaging sessions, particularly in the V3A.

2.5.2 Comparison with earlier studies of orientation representation

The perception of orientation is essential for reconstructing the 3D structure of an object. Previous studies have demonstrated that the parietal lobe plays an important role in orientation representation. Cues that support the depth discrimination of orientation include linear perspective and texture, velocity, and disparity gradients (Johnston and Passmore, 1994). These cues can be divided into the two following categories: monocular cues and binocular cues. In study of alert monkeys, Shikata et al. (1996) found that neurons in the lateral bank of the caudal part of the intraparietal sulcus (CIP) were sensitive to the orientation of a surface defined by disparity. Neurons from this region were also shown to be selective to binocular disparity gradients in monkeys (Sakata et al., 1998; Taira et al. 2000). Tuning to disparity-defined 3D surface orientation has also been found in TE regions (Janssen et al., 2000).

In our study of orientation representation of 3D surfaces defined by disparity, we first investigated

the orientation-selectivity of the neurons in ROIs using the "same-type stimuli horizontal–vertical surfaces classification." The accuracies of classification were higher than the baseline of statistical significance for all ROIs including the V1, V2, the ventral visual areas V3v and LOC, the dorsal visual areas V3d, V3A, KO, hMT+, and V7, and the IPS areas VIPS, POIPS, and DIPS. This finding is consistent with those of previous reports that disparity-selective neurons are widely distributed across visual cortical areas. To investigate the more generalized representation of horizontal–vertical orientation, we performed a "transfer horizontal–vertical classification of surfaces with different shapes at the same depth position." The results showed that from higher dorsal areas to the parietal areas V3A, V7, VIPS, and POIPS and ventral area LOC, the accuracies of classification were higher than chance level, suggesting that these areas are involved in the more generalized representations of orientation. The neurons in the CIP were previously shown to be sensitive to monocular texture gradient cues as well as to disparity, suggesting that they integrate texture and disparity gradient signals to construct a generalized representation of 3D surface orientation (Tsutsui et al. 2002). Ban and Welchman (2015) inferred the computational hierarchy that supports the estimations of slant using fMRI measurements and a series of generative models, and found that V3A was involved in slanted surface information representation by pooling disparity across space; the representation was largely unaffected by low-level stimulus changes, and V3A response showed a degree of tolerance across different depth positions. They concluded that V3A, which anatomically precedes CIP in macaques, is an earlier locus for processing the disparity signals of slant. This finding is consistent with our finding that the accuracy of classification in V3A was higher than the baseline of statistical significance. In addition, we found that the neurons in V7, VIPS, and POIPS may be related to the generalized representation of orientation. This is consistent with previous animal research showing CIP involvement in orientation representation through monocular cues such as texture gradient and perspective as well as binocular cues of binocular disparity, suggesting that representation in this area may be more generalized (Tsutsui et al., 2001; Tsutsui et al., 2002). Moreover, the relationship between human and monkey neuroanatomy is not exact; the human V7 and/or VIPS are considered to correspond to the macaque CIP (Orban et al., 2006a). Therefore, it is reasonable that the human areas V7 and VIPS were selective to orientation.

2.5.3 Other MVPA Analysis

Janssen et al. (2001) showed that neurons in the TEs represent differences in the second order disparities of 3D shapes. Moreover, they found that the neurons in TEs can code the orientation of curvature (i.e., vertical or horizontal). Therefore, one may believe that the representation of vertical

gradients (as in horizontally oriented cylinders) and that of horizontal gradients (as in vertically oriented cylinders) should be different. However, this may not necessarily be true. In the aforementioned study, the "selectivity for horizontal compared with vertical 3D shapes" was investigated, and the following combinations were verified: (1) the neurons selective for vertical but not for horizontal 3D shapes; (2) neurons selective for both directions of curvature; and (3) neurons selective for horizontal but not for vertical. A total of 104 neurons were tested, and the result showed that 16 of those neurons were selective for both vertical and horizontal 3D shape. In this regard, there may be neurons that are selective for both types of surfaces in some areas of the visual cortex. For a comprehensive investigation, we examined three additional types of classification in which horizontal surfaces and vertical surfaces were independently analyzed.

Three additional types of classification were also performed in our analysis. The first two types were related to the shape classification (Figure 2.12), whereas third type was related to the classification of the surface position in terms of depth (Figure 2.13). First, to identify the areas containing reliable information regarding the shape, we trained and tested the SVM to distinguish the shape using data of surfaces in the same orientation at both depth positions (hereafter denoted by A1). This classification is illustrated in Figure 2.12 (1.a and 1.b). The average accuracy for the two types of classification across eight participants is shown in Figure 2.14 (A). All the areas showed an average accuracy of classification that was higher than chance level for the two-class classification (50%). The V1; V2; dorsal areas V3d, V3A, V7, KO and hMT+; ventral areas V3v and LOC; and the parietal area VIPS exhibited an accuracy of classification that exceeded the baseline of statistical significance. Second, to investigate whether ROIs were involved in the generalized representation of the shape irrespective of the depth position of the surface, additional transfer shape classification was performed (hereafter denoted by A2). This classification is illustrated in Figure 2.12 (2.a, 2.b, 2.c and 2.d). We trained SVM using data of surfaces at the near position and tested using data of surfaces in the same orientation at the far position and vice versa. The average accuracy of the four types of classification across eight participants is shown in Figure 2.14 (B). The results showed that V1; V2; dorsal areas V3d, V3A and KO; and the parietal area VIPS exhibited an accuracy of classification that was statistically significant. Third, to identify the areas containing information regarding the depth position of the surface, an additional near–far classification was performed (hereafter denoted by A3). As illustrated in Figure 2.13 (1.a and 1.b), we trained and tested the SVM using data of surfaces for both the convex and concave shapes in the same orientation. The average accuracy of the two types of classification across eight participants are shown in Figure 2.14 (C). The results showed that accuracy of classification in all ROIs did not reach baseline of statistical significance.

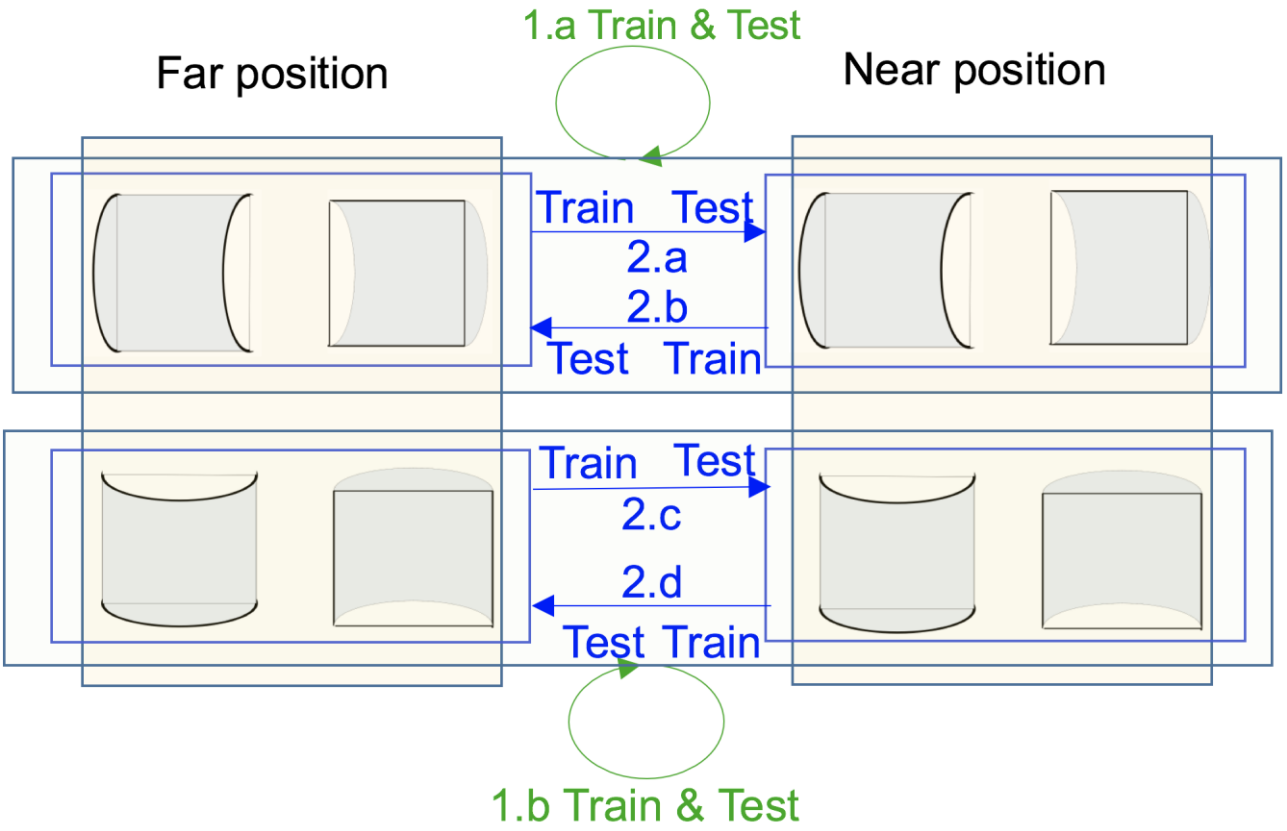

Figure 2.12 Illustration of the additional types of the shape classification performed. 1.a and 1.b represent the shape classification that trained and tested the SVM using data of surfaces with the same orientation at both depth positions (denoted by A1). 2.a, 2.b, 2.c, and 2.d represent the transfer shape classification that trained the SVM using data of surfaces at the near position and tested the SVM using data of surfaces at the far positon, and vice versa (denoted by A2).

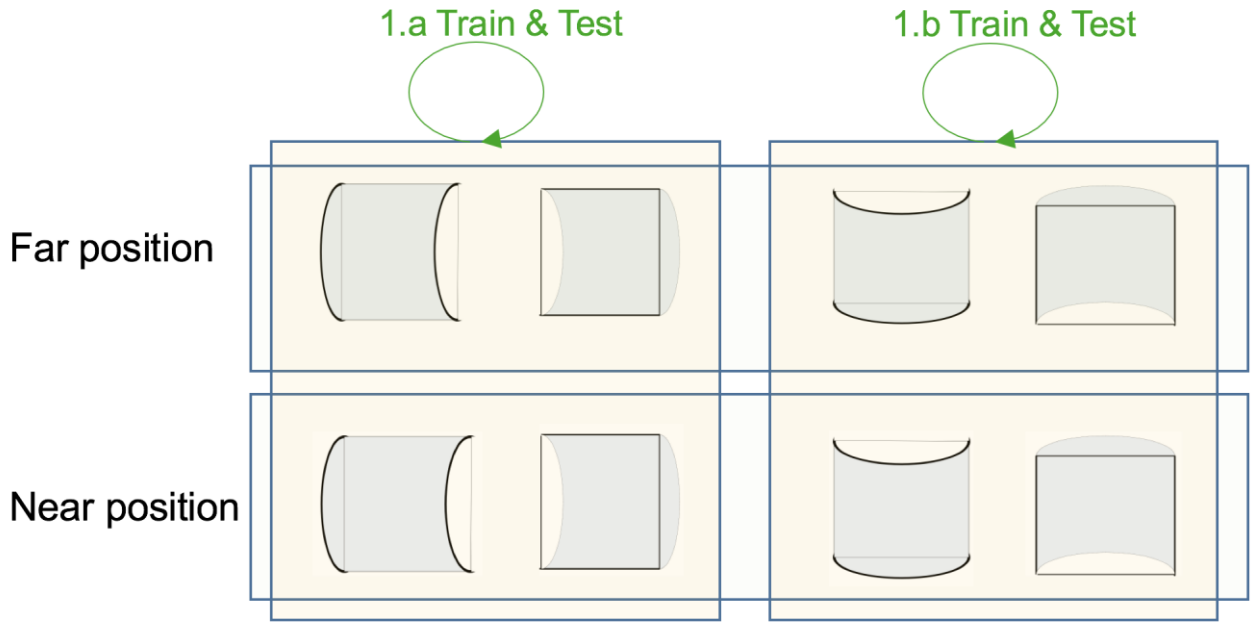

Figure 2.13 Illustration of the additional near–far classification performed. We trained and tested the SVM using data of both convex and concave surfaces with the same orientation (denoted by A3).

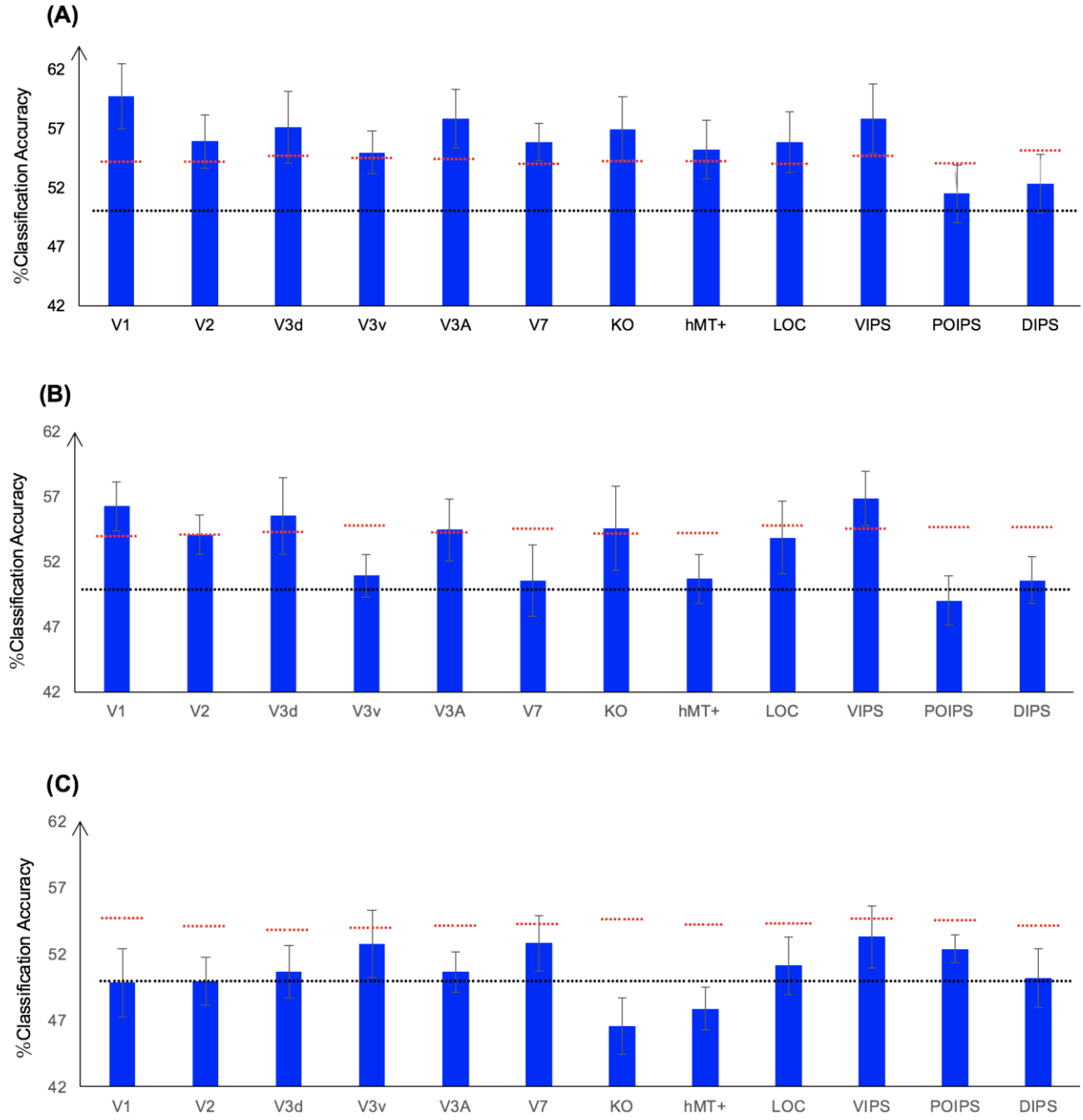

Figure 2.14 Accuracy of the additional types of classification. (A) The accuracy of classification for the additional classification A1. (B) The accuracy of classification for the additional classification A2. (C) The accuracy of classification for the additional classification A3. The red horizontal dotted lines indicate the baselines of statistical significance for the ROIs. The locations of these lines indicate the upper 99.6% percentile of the distribution of the accuracy of classification of the permutated data. The black horizontal line indicates the chance level for two-class classification.

Furthermore, the additional analysis showed the following.

(1) For classification A1, the additional results were similar (except for the accuracy, which was slightly higher in general) to those of the "same-type stimuli shape classification" in section 2.4.2.1 (hereafter, this type of classification is denoted by S1), in which both the training and testing data

were of the same depth position (i.e., near or far) and orientation (i.e., horizontal or vertical). Unlike in S1, in A1, V3v and hMT+ showed a classification accuracy that reached baseline of statistical significance. We reasoned that some additional information may render the shape classification easier in A1 versus S1. In classification type S1, all the convex and concave surfaces are in the same depth position. In classification type A1, the convex and concave surfaces in both the near and far depth positions are included. For convex and concave surfaces in the near depth position, if we add convex surface in the far depth position into the training and testing data, although the disparity difference between the convex surface in far depth position and concave surface in near depth position surrounding the foveal parts is smaller than the disparity difference between the convex surface in near depth position and concave surface in near depth position near the foveal part, the disparity in the peripheral part become larger. Therefore, it is possible that the accuracy of classification does not exhibit a considerable change or become slightly lower. Moreover, if we add concave surface in the far depth position into the training and testing data, the disparities between the convex surface in the near depth position and concave surface in far depth position is larger than disparities between convex surface in near depth position and concave surface in near position, particularly in the foveal part, and this larger disparity will probably increase the accuracy of classification. Therefore, the overall accuracy is increased.

(2) For classification A2, the results were similar to those reported for classification S1. However, the accuracy of classification in V7 and LOC was not significant. This observation can possibly be attributed to the representation of the shape V7 and LOC being more sensitive to the depth position. This can be inferred from the results of the "same-type stimuli near–far classification" (Figure 2.8). According to the results, V7, VIPS and POIPS showed high accuracies of classification, whereas other areas did not. Therefore, the V7 did not show a significant accuracy in this additional classification A2. For the LOC area, the accuracy of classification was not significant in near–far discrimination. However, Preston et al. (2008) showed LO (which is sub region of LOC) contained information regarding the depth position of planes depicted using a RDS. This may explain the lack of statistically significant accuracy for LOC area in this additional classification A2. The VIPS showed a significant accuracy in both classification A2 and near–far classification. This is probably because the VIPS contains a subgroup of neurons that are sensitive to the near–far depth position as well as neurons that are selective to the shape irrespective of the depth position.

(3) For classification A3, the results were inconsistent with those previously reported. A study showed that numerous areas from the early visual cortex areas to higher areas contained near–far information regarding the stereoscopic plain. Preston et al. (2008) found that numerous areas, such as

the retinotopic areas V1, V2, V3v, V4; higher ventral area LO; higher dorsal area hMT+, and IPS areas VIPS, POIPS and DIPS showed a significant accuracy of classification. The discrepancy between the results of the present study which showed no significant accuracy in all ROIs and those reported by previous study may be attributed to the following reasons. First, the stimuli used in the studies were different. We adopted curved surfaces in different depth positions, whereas Preston et al. adopted fronto-parallel flat planes in two different depth positions. It may be more difficult for the SVM to perform a near–far classification on curved than plain surfaces as the local disparity information is different within the curved surface. Second, we trained and tested the SVM using data of surfaces with a different shape. These surfaces with a different shape exhibited differences in local disparity, even at identical depth positions. Based on these reasons, we did not identify an ROI with a significant accuracy of classification.

2.5.4 Limitations

In our experiment, we adopted curved surfaces with different shapes and orientations. These surfaces were simulated in two different depth positions behind the fixation marker. For the surfaces in the same depth position, the surfaces with the same shape (i.e., convex or concave) and different orientations (i.e., horizontal or vertical) shared a similar average absolute disparity. In contrast, surfaces with a different shape (irrespective of orientation) at the same depth position exhibited differences in the average absolute disparity. For the surfaces in the different depth positions, the surfaces (irrespective of the shape) exhibited differences in the average absolute disparity.

For the shape classification, there were some limitations in our study: in the “transfer shape classification of surfaces at the same depth position,” we sought to investigate the generalized representation of the sign of surface curvature, irrespective of the orientation of surfaces. The differences in the average absolute disparity between surfaces with a different shape in the same depth position may have been used for this kind of shape classification. From this perspective, this may affect the conclusion of generalized representation of the shape of surfaces. However, most importantly, for the “transfer shape classification at different depth positions,” it was not possible to use the differences in average absolute disparity between surfaces with a different shape for classification. For this type of classification, the observed significantly high accuracy in the V3A cannot be interpreted as selectivity to average absolute disparity. However, it may only be interpreted

as a more generalized shape representation of surfaces. Therefore, it is reasonable to conclude that V3A was involved in the more generalized representation of the shape of surface, irrespective of its orientation and depth position.

For the orientation classification, no issues were noted due to the mismatch in the absolute disparity between surfaces with a different shape. In our study, we mainly performed two types of horizontal–vertical classification, namely the "same-type stimuli horizontal–vertical surfaces classification" and "transfer horizontal–vertical classification of surfaces with different shapes." The significantly high accuracy for these two types of classification could not have been caused by the mismatch in the absolute disparity between the convex and concave surfaces. However, we identified a few areas showing significant classification accuracy, indicating the involvement of a more generalized orientation representation.

In this investigation, we only selected convex and concave surfaces in two different orientations. For the reasons for the selection of these surfaces are discussed below.

First, the main focus of our study was to investigate the convex–concave shape and the horizontal–vertical representation of the stereoscopic surface in ROIs and to assess the involvement of generalized representation in each ROI. For the investigation of generalized representation of the shape and orientation, we required stimuli that exhibit both a shape and orientation features. Moreover, the shape feature and orientation feature can be easily and independently manipulated. Therefore, the convex and concave surfaces with two different orientations were considered. Other stimuli, such as "convex 'bumps' vs. concave 'dimples'" used by Dövencioğlu et al. (2013) only have the shape feature. Thus, they cannot be used to investigate the generalized shape representation or orientation representation.

Second, the surfaces we selected can be regarded as the element of a more complex surface (e.g., sinusoidal surface). We intended to initially investigate a simple and basic surface in our study first. Therefore, we selected convex and concave surfaces in two different orientations as our stimuli. We did not investigate a more complex index of surface, such as the Koenderink shape index (Koenderink et al., 2015), because it is more complicated, and it may be hard to show surfaces with differences large enough (e.g., two surfaces with a large curvature difference), which is not suitable for the MVPA method.

Another concern in this study is that the number of participants is relatively small compared with some previous MVPA studies. In the study by Murphy et al. (2013), data from 12 participants were used for analysis. However, there are studies that have included fewer or the same number of participants than ours. In a study by Horikawa and Kamitani (2017), five participants were recruited, whereas in the study by Preston et al. (2008), eight participants were recruited in the experiment. In our study, we assured reliability by performing "permutation tests." Permutation tests are widely used for significance testing in classification-based fMRI analyses (Ban and Welchman, 2015; Patten and Welchman, 2015). In our case, we randomized the correspondences between fMRI data and labels, and performed classification similar to that performed for normal, non-permutated data. This procedure was repeated 1,000 times to create a distribution of classification accuracies. We used the 99.6th percentile (one-tailed, 12 ROIs) as the baseline for statistical significance. This indicates that if a classification is higher than the baseline for statistical significance, it is extremely rare that the accuracy was obtained by chance in the eight participants. Therefore, we believe that our results are valid and reliable.

2.6 References of CHAPTER 2

Alizadeh S, Jamalabadi H, Schönauer M, Leibold, C, and Gais S (2017). Decoding cognitive concepts from neuroimaging data using multivariate pattern analysis. Neuroimage 159, 449-458.

Alizadeh AM, Van Dromme IC, and Janssen P (2018). Single-cell responses to three-dimensional structure in a functionally defined patch in macaque area TEO. J. Neurophysiol. 120, 2806-2818.

Anzai A, Chowdhury SA, and DeAngelis GC (2011). Coding of stereoscopic depth information in visual areas V3 and V3A. J. Neurosci. 31, 10270-10282.

Ban H, and Welchman AE (2015). fMRI analysis-by-synthesis reveals a dorsal hierarchy that extracts surface slant. J. Neurosci. 35, 9823-9835.

Chandrasekaran C, Canon V, Dahmen JC, Kourtzi Z, and Welchman AE (2007). Neural correlates of disparity-defined shape discrimination in the human brain. J. Neurophysiol. 97, 1553-1565.

Culham JC, Danckert SL, DeSouza JF, Gati JS, Menon RS, and Goodale MA (2003). Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. Exp. Brain Res. 153, 180-189.

Cumming BG, and Parker AJ (1997). Responses of primary visual cortical neurons to binocular

disparity without depth perception. Nature. 389, 280-283.

Cumming BG, and DeAngelis GC (2001). The physiology of stereopsis. Annu. Rev. Neurosci. 24, 203-238.

DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser J, Cox R, Miller D, and Neitz J (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. Proc. Natl. Acad. Sci. U. S. A. 93, 2382-2386.

Dövencioğlu D, Ban H, Schofield AJ, and Welchman AE (2013). Perceptual integration for qualitatively different 3-D cues in the human brain. J. Cogn. Neurosci. 25, 1527-1541.

Dupont P, De Bruyn B, Vandenberghe R, Rosier AM, Michiels J, Marhal G, Mortelmans L, and Orban GA (1997). The kinetic occipital region in human visual cortex. Cereb. Cortex 7, 283-292.

Fischl, B. (2012). FreeSurfer. NeuroImage 62, 774-781.

Fogassi L, Gallese V, Buccino G, Craighero L, Fadiga L, and Rizzolatti G (2001). Cortical mechanism for the visual guidance of hand grasping movements in the monkey: A reversible inactivation study. Brain 124, 571-586.

Freud E, Plaut DC, and Behrmann M (2016). 'What' is happening in the dorsal visual pathway. Trends Cogn. Sci. 20, 773-784.

Freud E, Ganel T, Shelef I, Hammer MD, Avidan G, and Behrmann M (2017). Three-dimensional representations of objects in dorsal cortex are dissociable from those in ventral cortex. Cereb. Cortex 27, 422-434.

Ganel T, and Goodale MA (2003). Visual control of action but not perception requires analytical processing of object shape. Nature 426, 664-667.

Georgieva S, Peeters R, Kolster H, Todd JT, and Orban GA (2009). The processing of three-dimensional shape from disparity in the human brain. J. Neurosci. 29, 727-742.

Georgieva SS, Todd JT, Peeters R, and Orban GA (2008). The extraction of 3D shape from texture and shading in the human brain. Cereb. Cortex 18: 2416-2438.

Goncalves NR, Ban H, Sánchez-Panchuelo RM, Francis ST, Schluppeck D, and Welchman AE (2015). 7 tesla FMRI reveals systematic functional organization for binocular disparity in dorsal visual cortex. J. Neurosci. 35, 3056-3072.

Goodale MA, Milner AD, Jakobson LS, and Carey DP (1991). A neurological dissociation between perceiving objects and grasping them. Nature 349, 154-156.

Goodale MA, and Milner AD (1992). Separate visual pathways for perception and action. Trends Neurosci. 15, 20-25.

Grossberg S, Srinisan K, and Yazdanbakhsh A (2011). On the road to invariant object recognition: how cortical area V2 transforms absolute to relative disparity during 3D vision. Neural Netw. 24, 686-692.

Horikawa T, and Kamitani Y (2017). Generic decoding of seen and imaged objects using hierarchical visual features. Nat. Commun. 8:15037.

Hubel DH, Wiesel TN, Yeagle EM, Lafer-Sousa R, and Conway BR (2015). Binocular stereoscopy in visual areas V-2, V-3, and V-3A of the macaque monkey. Cereb. Cortex 25, 959-971.

James TW, Culham J, Humphrey GK, Milner AD, and Goodale MA (2003). Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study. Brain 126, 2463-2457.

Janssen P, Vogels R, and Orban GA (1999). Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. Proc. Natl. Acad. Sci. U. S. A. 96, 8217–8222.

Janssen P, Vogels R, and Orban GA (2000). Three-dimensional shape coding in inferior temporal cortex. Neuron 27, 385-397.

Janssen P, Vogels R, Liu Y, and Orban GA (2001). Macaque inferior temporal neurons are selective for three-dimensional boundaries and surfaces. J. Neurosci. 21, 9419-9429.

Johnston A, and Passmore PJ (1994). Independent encoding of surface orientation and surface curvature. Vision Res. 34, 3005-3012.

Konen CS, and Kastner S (2008). Two hierarchically organized neural systems for object information in human visual cortex. Nat. Neurosci. 11, 224-231.

Koenderink JJ, and Doorn AJV (1992). Surface shape and curvature scales. Image Vis. Comput. 10, 557-564.

Kourtzi Z., and Kanawisher N (2000). Cortical regions involved in perceiving object shape. J. Neurosci. 20, 3310-3318.

Kourtzi Z, and Kanwisher N (2001). Representation of perceived object shape by the human lateral occipital complex. Science 293, 1506-1509.

Lehky SR, and Sereno AB (2007). Comparison of shape encoding in primate dorsal and ventral visual pathways. J. Neurophysiol. 97, 307-319.

Logothetis NK, Pauls J, Augath M, Trinath T, and Oeltermann A (2001). Neurophysiological

investigation of the basis of the fMRI signal. Nature 412, 150–157.

Marr D, and Poggio T (1976). Cooperative computation of stereo disparity. Science 194, 283-287.

Mishkin M, and Ungerleider LG (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Behav. Brain Res. 6, 57–77.

Monaco S, Chen Y, Medendorp WP, Crawford JD, Fiehler K, and Henriques DY (2014). Functional magnetic resonance imaging adaptation reveals the cortical networks for processing grasp-relevant object properties. Cereb. Cortex 24, 1540-1554.

Murphy AP, Ban H, and Welchman AE (2013). Integration of texture and disparity cues to surface slant in dorsal visual cortex. J. Neurophysiol. 110, 190–203.

Naganuma T, Nose I, Inoue K, Takemoto A, Katsuyama N, and Taira, M (2005). Information processing of geometrical features of a surface based on binocular disparity cues: an fMRI study. Neurosci. Res. 51, 147-155.

Neri P, Bridge H, and Heeger DJ (2004). Stereoscopic processing of absolute and relative disparity in human visual cortex. J. Neurophysiol. 92, 1880-1891.

Neri, P (2005). A stereoscopic look at visual cortex. J. Neurophysiol. 93, 1823-1826.

Nguyenkim JD, and DeAngelis GC (2003). Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. J. Neurosci. 23, 7117-7128.

Orban GA, Sunaert S, Todd JT, Van Hecke P, and Marchal G (1999). Human cortical regions involved in extracting depth from motion. Neuron 24: 929-940.

Orban GA, Claeys K, Nelissen K, Smans R, Sunaert S, Todd JT, Wardak C, Durand JB, and Vanduffel W (2006a). Mapping the parietal cortex of human and non-human primates. Neuropsychologia 44, 2647-2667.

Orban GA, Janssen P, and Vogels R (2006b). Extracting 3D structure from disparity. Trends Neursci. 29, 466-473.

Paradis AL, Cornilleau-Pérès V, Droulez J, Van De Moortele PE, Lobel E, Berthoz A, Le Bihan D, and Poline JB (2000). Visual perception of motion and 3-D structure from motion: an fMRI study. Cereb. Cortex 10, 772-783.

Parker AJ (2007). Binocular depth perception and the cerebral cortex. Nat. Rev. Neurosci. 8, 379-391.

Patten ML, and Welchman AE (2015). fMRI activity in posterior parietal cortex relates to the

perceptual use of binocular disparity for both signal-in-noise and feature difference tasks. PLoS One 10: e0140696.

Preston TJ, Li S, Kourtzi Z, and Welchman AE (2008). Multivoxel pattern selectivity for perceptual relevant binocular disparities in the human brain. J. Neurosci. 28, 11315-11327.

Preston TJ, Kourtze Z, and Welchman AE (2009). Adaptive estimation of three-dimensional structure in the human brain. J. Neurosci. 29, 1688-1698.

Roe AW, Parker AJ, Borm RT, and DeAngelis GC (2007). Disparity channels in early vision. J. Neurosci. 27, 11820-11831.

Rosenberg A, Cowan NJ, and Angelaki DE. (2013). The visual representation of 3D object orientation in parietal cortex. J. Neurosci. 33, 19352-19361.

Sakata H, Taira M, Murata A, and Mine S (1995). Neural mechanisms of visual guidance of hand action in the parietal cortex of the monkey. Cereb. Cortex 5, 429-438.

Sakata H, Taira M, Kusunoki M, Murata A, Tanaka Y, and Tsutsui K (1998). Neural coding of 3D features of objects for hand action in the parietal cortex of the monkey. Philos. Trans. R. Soc. Lond. 29, 1363-1373.

Schiller PH, Logothetis NK, and Charles ES (1990). Role of the color-opponent and broad-band channels in vision. Visual Neurosci. 29, 321-345.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, and Tooltell RB (1995). Borders of multiple visual areas in human revealed by functional magnetic resonance imaging. Science 268, 889-893.

Sereno ME, Trinath T, Augath M, and Logothetis NK (2002). Three-dimensional shape representation in monkey cortex. Neuron 14, 635-652.

Sim EJ, Helbig HB, Graf M, and Kiefer M (2015) When action observation facilitates visual perception: activation in visuo-motor areas contributes to object recognition. Cereb. Cortex 25, 2907-2918.

Shikata E, Tanaka Y, Nakamura H, Taira M, and Sakata H (1996). Selectivity of the parietal visual neurones in 3D orientation of surface of stereoscopic stimuli. NeuroReport 7, 2389-2394.

Shikata E, Hamzei F, Glauche V, Knab R, Dettmers C, Weiller C, and Büchel C (2001). Surface orientation discrimination activates caudal and anterior intraparietal sulcus in humans: an event-related fMRI study. J. Neurophysiol. 85, 1309-1314.

Taira M, Tsutsui KI, Jiang M, Yara K, and Sakata H (2000). Parietal neurons represent surface orientation from the gradient of binocular disparity. J. Neurophysiol. 83, 3140-3146.

Taira M, Nose I, Inoue K., and Tsutsui K (2001). Cortical areas related to attention to 3D surface structures based on shading: an fMRI study. NeuroImage 14: 959-966.

Takemura H, Rokem A, Winawer J, Yeatman JD, Wandell BA, and Pestilli F (2016). A major human white matter pathway between dorsal and ventral visual cortex. Cereb. Cortex 26, 2205–2214.

Thomas OM, Cumming BG, and Parker AJ (2002). A specialization for relative disparity in V2. Nat. Neurosci. 5, 472-478.

Tsutsui K, Jiang M, Yara K, Sakata H, and Taira M (2001). Integration of perspective and disparity cues in surface-orientation-selective neurons of area CIP. J. Neurophysiol. 86, 2856-2867.

Tsutsui K, Sakata H, Naganuma T, and Taira M (2002). Neural correlates for perception of 3D surface orientation from texture gradient. Science 298, 409-412.

Tootell RBH, Hadjikhani N, Hall EK, Marrett S, Vanduffel W, Vaughan JT, and Dale AM (1998). The retinotopy of visual spatial attention. Neuron 21, 1409-1422.

Tyler CW, Likova LT, Chen CC, Kontsevich LL, Schira MM, and Wade AR (2005). Extended concepts of occipital retinotopy. Curr. Med. Imaging Rev. 1, 319-329.

Uka T, and DeAngelis GC (2006). Linking neural representation to function in stereoscopic depth perception: roles of the middle temporal area in coarse versus fine disparity discrimination. J. Neurosci. 26, 6791-6802.

Vanduffel W, Fize D, Peuskens H, Denys K, Sunaert S, Todd JT, and Orban GA. (2002). Extracting 3D from motion: differences in human and monkey intraparietal cortex. Science 298, 413-415.

Van Dromme, I. C, Premereur, E., Verhoef, B. E, Vanduffel, W. and Janssen, P. (2016). Posterior parietal cortex drives inferotemporal activation during three-dimensional object vision. PLoS Biol. 14. doi: 10.1371/journal.pbio.1002445

Warnking J, Dojat M, Guérin-Dugué A, Delon-Martin C, Olympiesff S, Richard N, Chéhikian, A, and Segebarth C (2002). fMRI retinotopic mapping—step by step. NeuroImage 17, 1665-1683.

Woods RP, Grafton ST, Holmes CJ, Cherry SR, and Mazziotta JC (1998). Automated image registration: I. General methods and intrasubject, intramodality validation. J. Comput. Assist. Tomogr. 22: 139-152.

Zeki S (1978). Functional specialization in the visual cortex of the rhesus monkey. Nature 274, 423-

428.

Zeki S, Watson JD, Lueck CJ, Friston KJ, Kennard C, and Frackowiak RS (1991). A direct demonstration of functional specialization in human visual cortex. J. Neurosci. 11, 641-649.

Zeki S, Perry RJ, and Bartels A (2003). The processing of kinetic contours in the brain. Cereb. Cortex 13, 189-202.

CHAPTER 3

TOPIC 2: Shape representation from different depth cues

3.1 Abstract

The images projected onto the eyes are two-dimensional (2D), but we perceive a 3D world. A variety of depth cues are used to construct a 3D perception from these 2D images, including binocular disparity, perspective, texture, and motion parallax, but the underlying neural mechanisms are still not fully understood. To identify cortical regions involved in the processing of specific depth cues, we measured BOLD signals as participants viewed and classified 3D images of convex and concave shapes consisted of slanted planes with different image elements and depth cues. Three types of stimuli were used: (1) RDS with binocular disparity, (2) black−white dotted lines with perspective, and (3) black−white dotted lines with binocular disparity. The BOLD images were then classified by MVPA according to the activity patterns of early and higher-order visual cortical areas. To identify cortical visual areas selective for shape, we assessed the convex vs. concave discrimination accuracy of MVPA classifiers trained and tested using activity patterns evoked by the same stimulus type (same visual elements and depth cue). To identify cortical regions with shared neural activity patterns regardless of image type, convex vs. concave discrimination accuracy was assessed by transfer classification in which classifiers trained and tested using activity patterns evoked by convex or concave images with different visual depth cues (including different elements). Convex vs. concave classification accuracy using the same stimulus type was high in early visual areas and subregions of the intraparietal sulcus (IPS), whereas transfer convex vs. concave classification accuracy was high in subregions of the IPS. These results indicate that early visual areas process specific features of stimuli (including depth cues), whereas IPS regions perform more generalized processing of 3D shape independent of specific depth cue type.

3.2 Introduction

The ability to perceive the visual world in three dimensions is critical for survival. In humans, most activities of daily living depend on the ability to manipulate objects and self in 3D space. However, the images projected onto the retina are 2D and so must be reconstructed into an accurate 3D representation by the visual system. The information (cues) used for 3D reconstructions are termed depth cues. Depth cues can be classified into two types, binocular cues that depend on the difference in visual information acquired by the two retinas (binocular disparities) and monocular cues that can be acquired by only one eye such as perspective, texture, motion parallax, retinal image size, and interposition.

This 3D reconstruction involves a series of processing steps from the retina to higher-order visual cortices. According to the "two-stream" theory, visual information is processed progressively through the retina, thalamus, and primary visual cortex (V1) in occipital cortex and then divided into two anatomically and functionally separate streams, a dorsal stream from V1 to the parietal cortex and a ventral stream from V1 to the temporal cortex. Each stream processes visual information in a hierarchical manner, with each cortical region processing information based on the results (output) of lower-order regions (Zeki et al., 1978). Mishkin and Ungerleider (1982) first conceptualized this theory based on lesion research in non-human primates. The ventral stream has been termed the "what" stream as it processes information related to an object's identity (e.g., shape, size, color, and texture) whereas the dorsal stream has been termed the "where" stream as it processes visual information related to object location, movement, and spatial relationships. According to this perspective, 3D shape information should be processed by the ventral stream.

A revised "two-stream" theory was formulated by Goodale and Milner (1992). Rather than viewing both streams as contributing to conscious visual awareness, they argued that only the ventral stream contributes to conscious vision (known as the "perception" stream) while information in the dorsal pathway is used for the unconscious control of action, such as the movement of the body guided by visual input (and so is known as the "action" stream). According to this view, 3D shape information can also be processed in the dorsal stream for visually guided action. In fact, there is evidence that 3D shape information is processed both in the dorsal and ventral streams. In a macaque monkey study, Janssen et al. (1999) found that neurons in a subregion of inferior temporal (IT) cortex in the ventral stream are selective to 3D shape defined by disparity, while a human fMRI study by Taira et al. (2001) found that the intraparietal area in dorsal stream contributes to perception of 3D surface structure based on shading. Furthermore, an fMRI study by Freud et al. (2017) of patients with visual

object agnosia due to ventral cortex injury found that the intact dorsal cortex can produce 3D object representations without input from the ventral stream.

There is still no consensus on how 3D shape information is processed according to specific depth cues or on the specific functions of individual visual cortical areas in processing of 3D shape. Therefore, the first aim of the current study is to investigate which visual areas are responsive to 3D shape. To this end, we measured the neural activity patterns in visual pathways by BOLD fMRI (Logothetis et al., 2001) while subjects viewed 3D images of two simple shapes (a convex or concave) composed of different visual elements and with distinct depth cues. Then, MVPA was performed to classify the shapes using the BOLD signal patterns in various ROIs, including retinotopic visual cortices (V1, V2, V3d, V3v, V3A, and V7), higher ventral cortex LOC, higher dorsal area hMT+, KO, and IPS areas VIPS, POIPS, and DIPS. The classifier was trained and tested using activity patterns in response to the images (convex or concave) with the same type of cue (perspective or binocular disparity), a condition termed "same cue type convex vs. concave classification." If classification accuracy using the signal from a given ROI is higher than the statistical significance level, we can conclude that neural activity pattern in the ROI includes information which support classifying the 3D shape from a specific cue.

A variety of depth cues contribute to 3D shape perception, and many models have been proposed to explain how different depth cues contribute to 3D shape processing at the neural level. Maloney and Landy (1989) proposed a simple statistical framework for combining depth estimates from consistent depth cues in which information from different cues is processed independently by different modules and then fused into a single depth estimate about each point of the scene. They also assumed that the assigned weights of different estimates are variable and that the combination is linear (additive). In contrast, Clark and Yuille (1990) distinguished between "strong fusion" and "weak fusion". In strong fusion, information from different cues interacts and is processed cooperatively to yield a single depth estimate, whereas in weak fusion information from different cues is independent and the final estimate is obtained by combining the individual information. However, the functions of specific cortical structures in processing different depth cues and their potential interactions are unknown. There are two broad possibilities: (1) regional neural activity patterns differ depending on cue type(s) or (2) shape information from different depth cues is fused so that neural activity patterns are similar among some regions irrespective of cue type. Therefore, the second aim of this study is to investigate

whether the neural activity patterns differ during processing of a given 3D shape with distinct depth cues or if there are common patterns among some regions independent of cue type. We examined this question using MVPA to classify BOLD data for each ROI obtained during processing of images with cue types different from that of the training set, a condition termed "transfer convex vs. concave stimuli classification". Classification accuracy higher than statistical significance level in a given ROI would suggest that there are common neural activity patterns for processing shapes using different depth cues (i.e., the activity pattern in a given ROI during processing of the same shape is independent of the depth cue type).

We predicted that many early- and some mid- to higher-level visual areas would show high classification accuracy for the "same cue type convex vs. concave classification" condition (i.e., dependence on cue type) for several reasons. First, early visual areas are believed to process simple attributes; for example, neurons in V1 are selective to orientation, binocular disparity, and motion direction (Hubel and Wiesel, 1962). For the convex vs. concave classification task, these local features are very different. In addition, shape processing occurs in some mid- to higher-order dorsal and ventral stream regions (Todd, 2004). Conversely, we also expected that some higher dorsal visual areas, especially parietal areas, would show high classification accuracy for transfer convex vs. concave classification (i.e., different cue types for training and testing) because these regions process more complex attributes of objects basing on previous steps. Indeed, there is evidence that certain parietal areas are involved in processing 3D shapes defined by different cues. A fMRI study by Durand et al. (2009) found that the anterior IPS regions are involved in the processing of 3D shape defined by disparity, and Nelissen et al. (2009) found that processing of 3D shape based on texture involves both ventral and parietal regions, with strongest activation observed in the caudal intraparietal region (CIP) and decreasing strength toward the anterior IPS. In addition, dorsal visual stream is believed to be involved in the unconscious control of action. For instance, the intraparietal cortex is involved in the control of visually guided action, like reach-to-grasp, which requires on the rapid extraction of 3D shape information. It is possible that neurons in some parietal areas will show a common neural activity pattern when 3D shapes defined by different cues are presented. Therefore, we expected that higher areas especially parietal areas would show high accuracy for transfer convex vs. concave classification.

3.3 Methodology

3.3.1 Participants

Nine participants were recruited for the fMRI experiments, seven males and two females. One male was left-handed, whereas all other participants were right-handed. All participants had normal or corrected-to-normal vision. None had any history of mental illness or neurological disease. Their ages ranged from 20 to 34 years (mean ± SD, 23.9 ± 3.9 years). Participants were remunerated for their participation.

The study protocol was approved by the Human Research Ethics Committee of the Kochi University of Technology. Written informed consent was obtained from all participants in accordance with the Declaration of Helsinki.

3.3.2 Stimuli

Stereoscopic stimuli were presented using a PROPixx projector with a 3D circular polarized filter placed in front. The stimuli were projected onto a translucent screen inside the bore of the fMRI magnet. The participants wore polarized glasses and viewed the images through a slanted mirror (angled at 45°) above the head coil. The optical distance from the midpoint of the two eyes to the screen was 71 cm. The screen resolution was set at 1920 × 1080 pixe ls and the refresh rate was 480 Hz.

The 3D stimuli were generated using Psychotoolbox 3 in MATLAB (MathWorks, Natick, MA, USA). The images depicted a shape that was either convex or concave consisted of two slanted planes using different visual components and depth cues: (1) RDS with disparity, (2) black–white dotted lines with disparity, or (3) black–white dotted lines with perspective (Figure 3.1). Different depth cues with similar elements (stimulus 2 and 3) were used to examine whether cues can induce cue-specific or common neural activity patterns. Similarly, images with different elements and same depth cues (stimulus 1 and 3) were used to assess activity related to visual elements. Images were presented on a mid-gray background rectangular area of 31.7° × 18.2°. A fixation marker comprised of a hollow square with a side length of 0.7° and horizontal and vertical nonius lines of 0.5° was shown at the center of the screen to help participants maintain eye vergence.

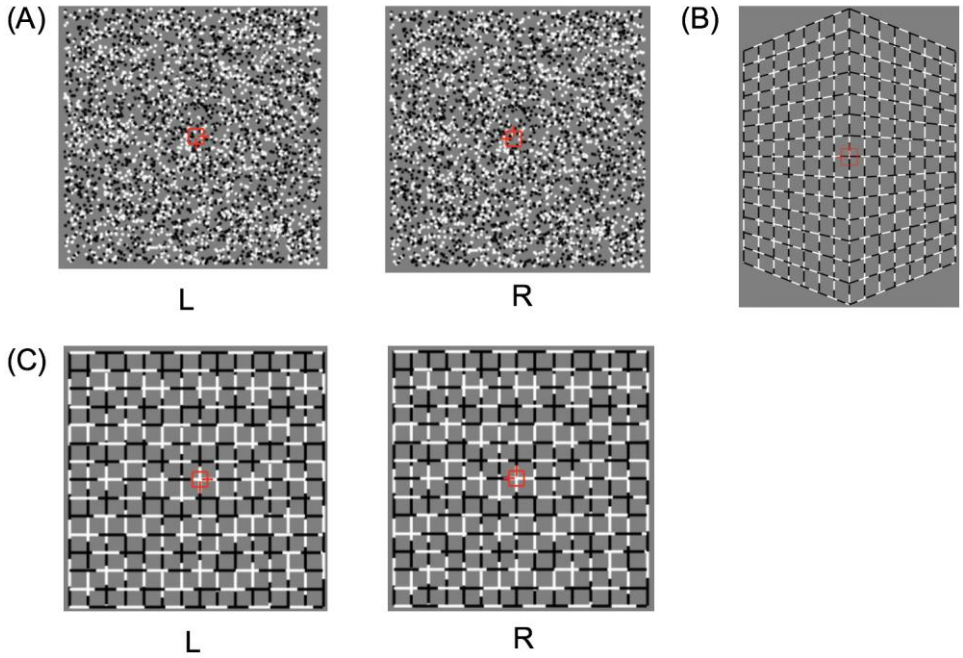

Figure 3.1 Sample visual stimuli used to generate 3D images of convex or concave shape consisted of planes (only those generating the convex shape are shown). (A) RDS. (B) Black−white dotted lines with perspective. (C) Black−white dotted lines with disparity.

For the RDS, the stimulus for each eye covered an area of 14.6° × 14.6°. The density of the stereogram was 289 dots/deg^2 and all dots were of the same size (0.2° in diameter). The relative disparity between the peak and sides of the convex shape was 0.4° and that of the concave shape was 0.2°. The angle of the two slanted planes forming the convex shape was 113° and that of the planes forming the concave shape was 121°. Both convex and concave peaks were at the fixation marker plane. The concave shape was nearer than the fixation marker plane with crossed disparity, whereas the convex shape was farther than the fixation marker plane with uncrossed disparity. Figure 3.2 presents a top view of the virtual 3D images.

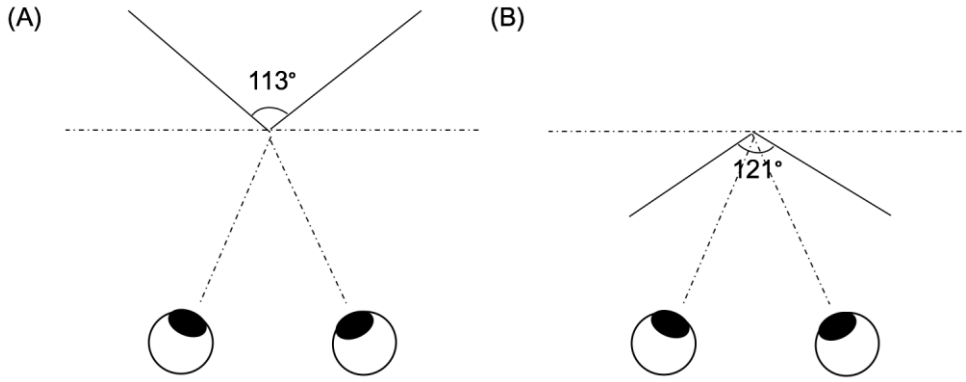

Figure 3.2 Top views of the virtual 3D images (using the RDS stimuli as the example). (A) The convex shape. (B) The concave shape.

For the stereo images composed of black–white dotted lines with disparity, each monocular component covered an area of 14.6° ×14.6°. Lines consisted of black and white sublines of random length. We used dotted lines because it is easier for participants to perceive convex and concave shapes compared to solid black or white lines with disparity. In the latter case, participants may not be able to correctly match the corresponding lines in the left and right eye. The angle of the two slanted planes forming these convex and concave images differed among participants. To determine the optimal angle for 3D perception, participants first performed a task to adjust the angle defined by dotted lines with disparity to that defined by RDS before the fMRI scan. In this procedure, the reference shape defined by RDS and the shape for adjusting dotted lines with disparity were shown in sequence, and the participants adjusted the angle of the latter image by pressing the corresponding buttons on a keypad. This task was repeated twice by each participant, and the average disparity was used for the main experiment.

For the images composed of black–white dotted lines with perspective, we used the same types of lines as used for images with disparity. The participants also performed a similar angle adjustment task so that the angle of shape defined by perspective appeared the same as the angle of shape defined by RDS. The width of the stimuli was 14.6° and height differed depending on the angle adjustment by the participants.

3.3.3 Experimental design

Convex and concave shapes composed of different elements or cues were presented: (1) RDS, (2) black–white dotted lines with disparity, (3) black–white dotted lines with perspective. Thus, there were six different stimuli (3 types ×2 shapes), which were presented in a block design (Fig ure 3.3). In order to avoid adaptation and to maintain neuronal activation, the stimuli in each block were flashed on and off at 1 Hz (each on and off period lasting 0.5 s). The random dots or black–white dotted lines were regenerated on every presentation. In each block, one of the 6 stimulus conditions was shown for 15 s. After each block, the participant was required to judge whether the shape was convex or concave by pressing the corresponding button on a keyboard within 3 s. After each judgment, there was a 6 s fixation period (the last fixation block was 12 s). All six stimulus conditions were presented in random order (12 blocks). Each run of began with a 12 s fixation period. The total time taken for one run is calculated as follows: (1) the time of all stimulus blocks is 15 × 12 = 180 s; (2) the time of all blocks for judgment is 3 × 12 = 36 s; and (3) the time of fixation blocks is 12 + 6 ×(12 − 1) + 12 = 90 s. Therefore, each run lasted a total of 180 + 36 + 90 = 306 s.

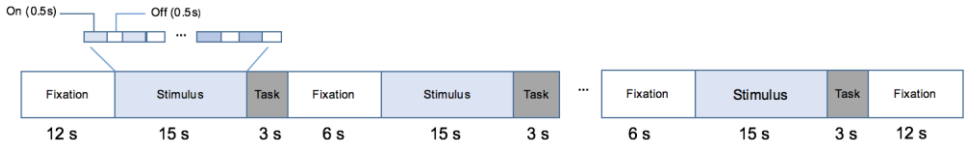

Figure 3.3 The block design of our experiment.

Participants were instructed to observe the fixation marker during the scan, and try their best to reduce head movement as much as possible over the entire session. If head movement was too large during one run (translational movement > 2 mm or rotation > 2°), the run was excluded from analysis. In general, all participants were able to maintain head stillness and at least 6 runs of EPI data were acquired for analysis from each participant.

3.3.4 Data acquisition

Functional MRI data acquisition was performed at the Brain Communication Research Centre of the Kochi University of Technology using a 3 Tesla Siemens MRI scanner with a 24-channel multi-phase array head coil. A pair of foam pads was used to help participants keep their head still during the experiment. To construct a flattened cortical surface with preserved relative dimensions, a high resolution (1 ×1 ×1 mm) T1 -weighted anatomical image was obtained for each participant. During the functional scans, we measured BOLD signals using an EPI sequence from an area of 35 slices that covered the visual cortex, posterior parietal cortex, and posterior temporal cortex. The acquisition parameters of the EPI sequence were as follows: 102 volumes per run; echo time (TE) of 30 ms; repetition time (TR) of 3000 ms; slice thickness of 3 mm; descending slice acquisition order. For each session, we also measured a T2-weighted structural image with the same number of slices and covering the same area as the corresponding EPI data. This structural image was used as the reference for 3D motion correction of EPI data and for co-registration between the T1-weighted anatomical image and EPI images. After co-registration, both the anatomical image and the EPI images were transformed into Talairach space.

To define the ROIs, separate fMRI measurements were acquired from all participants before the main experiment using stimuli shown to active specific cortical regions in previous studies. A rotating wedge and expanding rings were used to define the retinotopically organized visual areas V1, V2, V3d, V3v, and V3A (Sereno et al., 1995; DeYoe et al., 1996; Warnking et al., 2002). The V7 was defined as the area dorsal and anterior to V3A with lower visual field quadrant representation (Tootel et al., 1998; Tyler et al., 2005). In addition, the higher dorsal areas hMT+ and

KO, the higher ventral area LOC, and areas along the IPS (VIPS, POIPS, and DIPS) were localized using standard procedures. Briefly, hMT+ was defined as an area of voxels in the lateral temporal cortex exhibiting significantly higher ($p < 10^{-4}$) activation to dots that moved outward and inward coherently compared to static dots (Zeki et al., 1991), while KO was defined as an area of voxels showing significantly higher ($p < 10^{-4}$) activation in response to contours defined by motion compared to transparent motion of black and white dots (Dupont et al., 1997; Zeki et al., 2003). LOC was defined as an area of voxels in the lateral occipito-temporal cortex demonstrating significantly higher ($p < 10^{-4}$) activation in response to intact images of objects compared to the corresponding scrambled image (Kourtzi et al., 2000; Kourtzi et al., 2001). Finally, individual areas along IPS (VIPS, POIPS, and DIPS) were localized using nine connected lines as stimuli. These areas were identified by contrasting activity to 3D shapes produced by rotating the lines in depth compared to movement along a frontoparallel plane (Vanduffel et al., 2002).

3.3.5 Data analysis

3.3.5.1 Pre-processing

Data processing and analysis were conducted using Freesurfer (Fischl, 2012), BrainVoyager QX (Version 2.8.4.2645, 64-bit; BrainInnovation, Maastricht, The Netherlands), MATLAB R2014a, and SPSS Statistics 23 (IBM Inc.). The scalp was removed from the T1-weighted 3D anatomical images of each participant and white matter (WM) separated from the other components using Freesurfer. The WM component was then used as the mask to segment WM from gray matter (GM) using Brain Voyager QX. The extracted brain was transformed into Talairach space using BrainVoyager QX. The flattened cortical surface was generated to visualize functional maps and delineate ROIs by the following series of operations: segmenting the brain along the GM/WM boundary, inflating the segmented GM, cutting the GM along the calcarine sulcus, and flattening. The ROIs were then used for MVPA. Slice scan time correction and 3D motion correction of EPI data were performed using a T2-weighted reference image obtained at the beginning of each session (Wood et al., 1998). Temporal filtering was then applied on these EPI data. Spatial smoothing was not performed. Co-registration was conducted to align the functional EPI data to the anatomical T1-weighted data. Finally, functional EPI data were transformed into Talairach space.

3.3.5.2 ROI-based MVPA

The high sensitivity of MVPA allows for the detection of subtle differences between conditions of

interest and comparing neural patterns between cortical ROIs of human brain. In this study, we performed MVPA for the ROIs in MATLAB. The linear SVM implemented in MATLAB was used as the classifier for MVPA. Two main types of classification were performed. In this first classification type, SVM was trained and tested using activity patterns evoked by shapes with the same type of depth cue (e.g., using data from the shapes by RDS for both training and testing). This classification type was used to assess whether reliable neural activity patterns are induced by convex and concave shapes with specific cue types. The second classification type, transfer convex vs. concave classification, used activity patterns evoked by images with different cues for training and testing (e.g., using activity evoked by the shapes from RDS for training and the shapes from perspective for testing). The main purpose of this second type of classification was to assess whether certain areas within the visual cortex exhibit common neuronal activity patterns in response to convex or concave shapes irrespective of cue. Figure 3.4 illustrates these classifications.

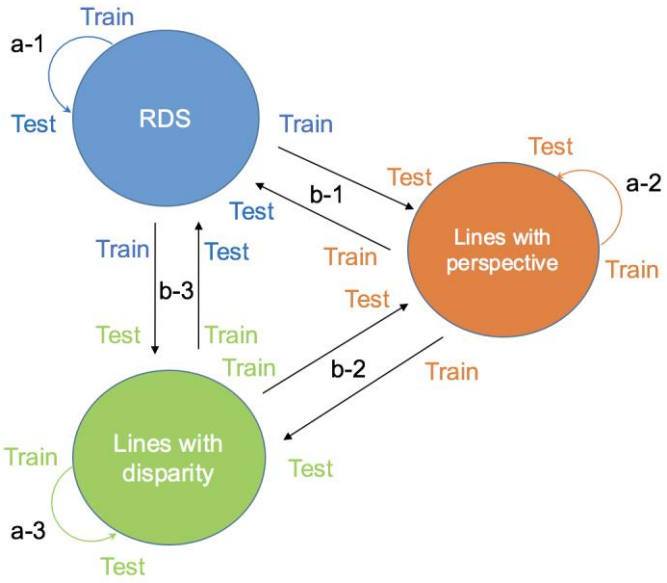

Figure 3.4 Illustration of the classifications performed. a-1, a-2, and a-3 indicate classification of convex vs. concave 3D images generated with the same depth cue; b-1, b-2, and b-3 indicate classification of convex vs. concave 3D images generated with different depth cues.

The MVPA procedure was conducted as follows. All fMRI time series were shifted two volumes (6 s) to account for the hemodynamic delay of the BOLD signal. For each ROI, voxels were selected from both the left and right hemispheres. These voxels were sorted from large to small response magnitude compared to the eye-fixation baseline condition according to the t-statistic. We selected the top 250 voxels from each ROI for classification. If the total number of voxels was less than 250 for a given ROI, the largest available number of voxels was used for the classification.

To estimate the response amplitude of each voxel of ROI within a stimulus block, we calculated the difference between the average BOLD signal of the first three volumes (after shifting by 2 volumes)

after stimulus onset (denoted avg1) and the average BOLD signal value of the last two volumes (after shifting by 2 volumes) before stimulus onset (denoted avg2) or avg1–avg2. Finally, all difference values were transformed into z-scores and these z-scores were used as the input for SVM training and testing.

To evaluate the performance of MVPA classification, the leave-one-run out cross-classification method was used to partition the EPI data into training and testing datasets consisting of different combinations of data from individual subject runs. For each individual subject ROI, the accuracies of all cross-classifications were averaged. Finally, the averaged accuracy of each individual ROI was averaged across participants.

To assess whether the classification accuracy was reliable, permutation testing was performed to estimate the baseline of statistical significance. Classification was performed using randomly permutated fMRI patterns for all ROIs (i.e., the correspondences between the fMRI data and the class labels were randomized) and classification was performed in the same way as for normal non-permutated data. We repeated this procedure 1000 times to generate a distribution of classification accuracies. The 99.6th percentile (one-tailed, 12 ROIs) was used as the baseline for statistical significance.

3.4 Result

3.4.1 Classification of convex vs. concave 3D images with the same depth cue

To investigate whether neurons in ROIs are selective to convex or concave 3D images generated from RDS, black−white lines with perspective, and black−white lines with disparity, we measured BOLD signals from visual cortex and then subjected the ROI data to MPVA using the same image type for training and testing (e.g., training using RDS data and testing using RDS data).

3.3.1.1 Classification of convex vs. concave image composed of random dots

This classification process is illustrated Figure 3.4 a-1 and the results for all ROIs are shown in Figure 3.5(A). Almost all ROIs showed classification accuracy (convex vs. concave) greater than chance for this binary classification (50%). Among these ROIs, V1, dorsal areas V3d, and V3A demonstrated classification accuracy higher than the baseline of statistical significance, and DIPS showed classification accuracy only slightly below the baseline of statistical significance.

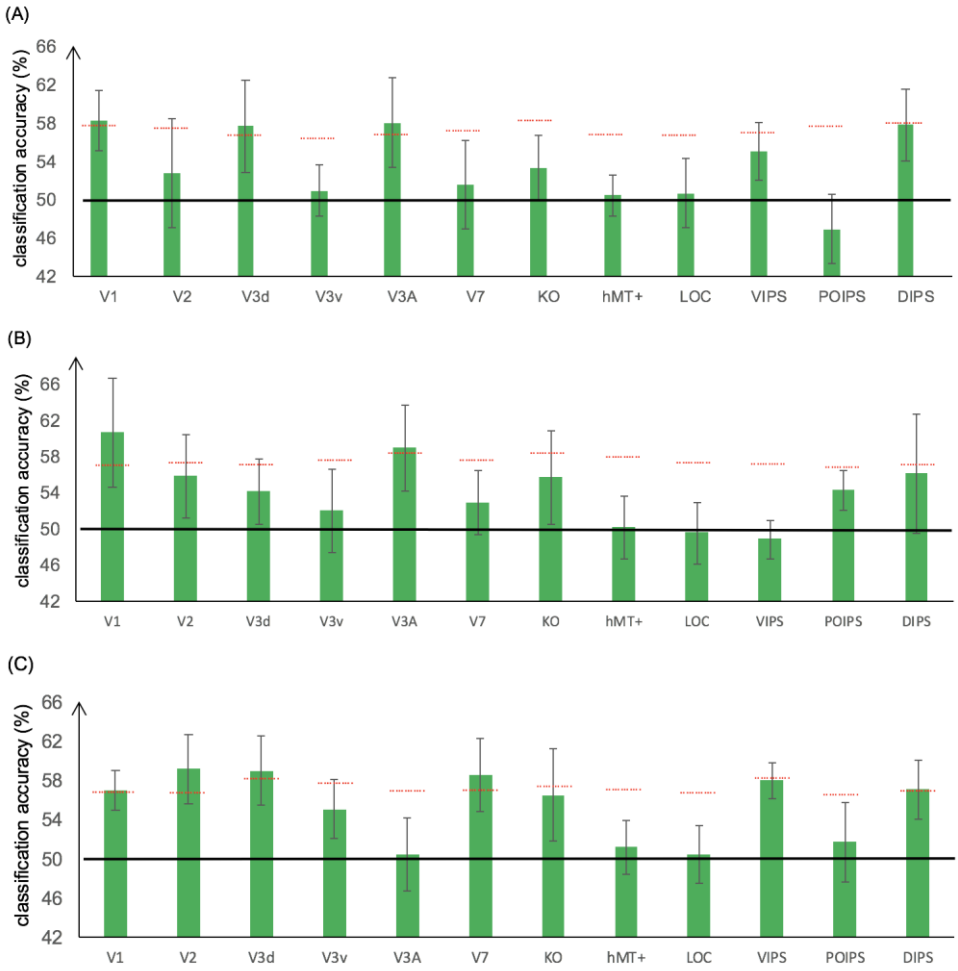

Figure 3.5 Classification accuracies for convex vs. concave 3D images generated with the same depth cue. (A) RDS data. (B) Lines with perspective. (C) Lines with disparity. The red horizontal dotted lines indicate the baseline of statistical significance for each ROI. The locations of these lines indicate the upper 99.6% percentile of the classification accuracy distribution for permutated data. The black horizontal line indicates the chance level for the binary convex vs. concave classification (50%). The error bars depict the standard error of the mean across subjects (n = 9).

3.4.1.2 Classification of convex vs. concave images composed of lines with perspective

This classification process is illustrated in Figure 3.4 a-2 and the results for all ROIs are shown in Figure 3.5 (B). Again, many ROIs showed classification accuracy higher than chance (50%). Among these ROIs, V1 and dorsal areas V3A demonstrated classification accuracy higher than the baseline of statistical significance.

3.4.1.3 Classification of convex vs. concave images composed of lines with disparity

This classification process is illustrated in Figure 3.4 a-3 and the results for all ROIs are shown in Figure 3.5(C). Many ROIs showed classification accuracy higher than chance (50%), and the ROIs for V1, V2, dorsal areas of V3d and V7, and intraparietal area DIPS demonstrated classification

accuracy higher than the baseline of statistical significance.

3.4.2 Classification of convex vs. concave 3D images with different depth cues or different visual elements

In the second type of classification, SVM was trained and tested using different depth cues (including different visual elements condition) to investigate whether certain ROIs exhibit a common neural activity pattern in response to a given image (concave or convex) independent of the specific depth cue (e.g., training using images generated from black−white lines with perspective and testing using images generated from black−white lines with disparity).

3.4.2.1 RDS for training and lines with perspective for testing and vice versa

In this classification task, we trained SVM using the 3D shapes generated by RDS and tested SVM using shapes generated by black−white dotted lines with perspective and vice versa. The classification process is illustrated in Figure 3.4 b-1 and results are shown in Figure 3.6(A). Only intraparietal area DIPS showed classification accuracy higher than the baseline of statistical significance.

3.4.2.2 Lines with perspective for training and lines with disparity for testing and vice versa

In this classification task, we trained SVM using 3D shapes generated by black−white dotted lines with perspective and tested SVM using shapes generated by black−white dotted lines with disparity (and vice versa). This classification is shown in Figure 3.4b-2 and the results are shown in Fig 3.6 (B). Only intraparietal areas POIPS and DIPS showed classification accuracy higher than the baseline of statistical significance.

3.4.2.3 RDS and lines with disparity RDS for training and lines with disparity for testing and vice versa

We also examined classification accuracy for images generated using different visual elements but the same depth cue (RDS for training and lines with disparity for testing and vice versa). This type of classification is illustrated in Fig 3.4b-3 and results are shown in Fig 3.6(C). The dorsal areas V3A and, V7 as well as IPS area DIPS demonstrated classification accuracy higher than the baseline of

statistical significance.

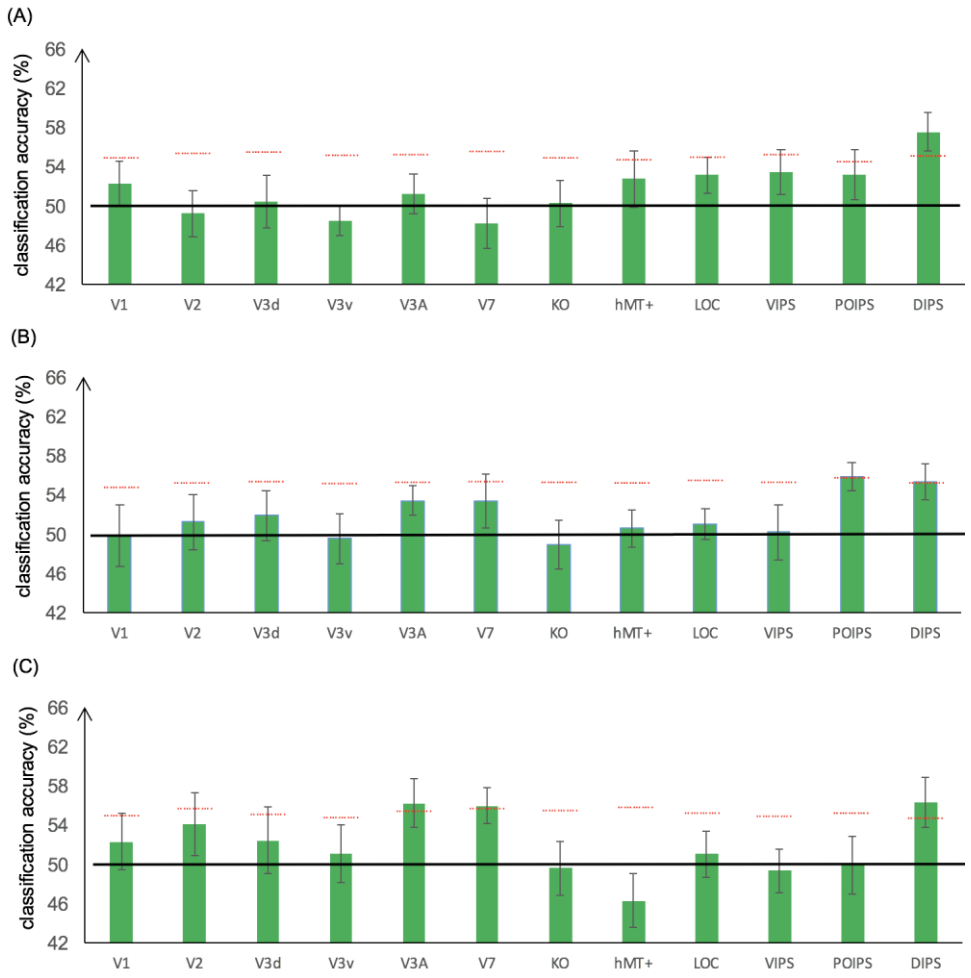

Figure 3.6 Classification accuracies for convex vs. concave 3D images generated with different depth cues (including different elements condition). (A) RDS and lines with perspective. (B) Lines with perspective and lines with disparity. (C) RDS and lines with disparity. The red horizontal dotted lines indicate the baseline of statistical significance for each ROI. The locations of these lines indicate the upper 99.6% percentile of the classification accuracy distribution for permutated data. The black horizontal line indicates the chance level. The error bars depict the standard error of the mean across subjects (n = 9).

3.5 Discussion

In the present study, we examined the processing of different visual elements (dots and lines) and depth cues (disparity and perspective) by different subregions of the visual cortex using fMRI and MVPA to elucidate the mechanisms of 3D image processing. To examine processing of visual elements, measurements were conducted using similar images (a convex or concave shape) generated by a RDS and by black−white dotted lines, both with binocular disparity as the depth cue. Alternatively, processing of depth cues was examined by measuring responses to similar images generated by the same elements (black−white dotted lines) but with different depth cues (disparity or

perspective). The major findings can be summarized as follows. (I) When the depth cue (including element) was the same, regions with activity showing significant classification accuracy by SVM learning were mainly located in lower- and mid-level visual areas, and the higher-order DIPS region also demonstrated significant or near significant classification accuracy. (II) When the depth cues (or elements) were different, the DIPS area and other higher-order areas demonstrated high accuracy. These results suggest that in general, lower level visual areas generate unique activity patterns in response to the same image with different depth cues, while some of higher-level regions generate similar activity patterns in response to the same image with different depth cues.

3.5.1 Differential response of various low- and high-level cortices to depth cues

It is believed that the visual system is split into two separate pathways after primary visual cortex (V1), a dorsal pathway to parietal cortex and a ventral pathway to IT cortex. Visual information is processed progressively in a hierarchical manner from lower- to higher-order visual areas. Lower-order visual areas in each pathway mainly process simple attributes, such as motion direction, orientation, and speed, whereas the visual attributes processed by higher-order visual areas of each pathway are much more complex, culminating in recognition of objects and complex shapes. We hypothesized that lower-order visual areas would show high accuracy in the same cue (including element) type convex vs. concave classification condition, and that some higher-order visual areas would show high accuracy in both the same cue (and element) type and different cue (including different element) type (transfer) convex vs. concave classification conditions. In general, the MVPA results are in line with our expectations.

(1) The classification accuracy of V1 was above the baseline of statistical significance for all convex vs. concave classifications with the same cue (and element) type. This type of classification includes three subtypes depending on the stimuli used. For the same cue type (lines with perspective) convex vs. concave classification condition, the convex and concave shapes are distinguished by line orientation, and V1 neurons are highly sensitive to lines of different orientation (Hubel and Wiesel, 1959). For the same cue type (RDS or lines with disparity) convex vs. concave classification, it is possible that the high accuracy was dependent on the selectivity of V1 neurons to the sign of disparity gradient: the disparity sign is different for concave and convex forms, and V1 neurons are highly sensitive to the sign of binocular disparity (Preston et al., 2008). In addition, the convex shapes provided information of uncrossed disparity and the concave shapes provided information of crossed disparity, which may differentially activate populations of V1 neurons. It is also possible that the high classification accuracy of V1 for these two conditions may have been aided by the

difference in image depth (near or far to the fixation marker image plane).

(2) The classification accuracy of V3d was high for the same cue (disparity) convex vs. concave condition (same elements: dots or lines), possibly because V3d neurons are also sensitive to binocular disparity. In an fMRI study by Preston et al. (2008), neurons in V3d demonstrated selectivity to absolute disparity, while Chandrasekaran et al. (2006) reported that fMRI activity in area V3d is related to performance in binocular disparity-defined shape judgments. Therefore, the high classification accuracy in these two conditions is consistent with the response properties of V3d neurons. Conversely, classification accuracy of V3d activity was below significance for the convex vs. concave classification of images generated by lines with perspective, suggesting that V3d neurons are not selective to perspective or that perspective-selective neurons are not clustered densely enough (and thus the local signal is not strong enough) for SVM classification. Classification accuracy was also not significant for the transfer convex vs. concave classification conditions using different depth cues. We speculate that this classification requires high-order processing (greater integration), while V3d is an early visual area in the dorsal pathway. At a lower level of processing, differences in training and testing conditions would induce different activity patterns for the same shape type, resulting in poor classification accuracy.

(3) Classification accuracy of the dorsal area V3A was significant for the same cue convex vs. concave conditions RDS and lines with perspective as well as for the transfer condition RDS/lines with disparity. These findings may be explained by the selectively of V3A neurons for disparity (Goncalves et al., 2015). Moreover, our previous fMRI study in topic 1 showed that V3A is selective to stereoscopic convex–concave shapes. The disparity patterns between the convex and concave shapes defined by RDS were different, so the high accuracy for the RDS convex vs. concave classification condition is consistent with known response properties of V3A neurons. In addition, there is evidence that basic perspective processing can be performed in V3A (Welchman et al., 2005), thereby accounting for the high accuracy for lines with perspective convex vs. concave classification. While both RDS and lines with disparity create 3D convex and concave shapes by binocular disparity, classification accuracy was high for the RDS condition but around chance for the lines with disparity condition. As V3A neurons are selective for shape from disparity, it is reasonable that classification accuracy would be high for RDS. However, the shapes generated by lines of disparity were depicted by conflicting cues: the disparity produced the convex and concave shapes, whereas the horizontal and vertical parallel lines created the flat plane. As mentioned, V3A neurons are also involved in the processing of perspective. Therefore, V3A accuracy in the lines with disparity condition was near chance level.

The V3A classification accuracy for the RDS/lines with disparity transfer condition was also higher than the baseline of statistical significance, consistent with our study in topic 1 in which we found that V3A neurons showed a common activity pattern to convex and concave shapes defined by disparity in two different orientations and depth positions. It is thus possible that area V3A is involved in a generalized representation of shape defined by binocular disparity and do not depend element type (random dots vs. lines).

(4) IPS areas

The classification accuracy of the DIPS was around the baseline of statistical significance for the same cue (and element) condition and higher than the baseline of statistical significance for all transfer (different depth cue or element) conditions. Mid- and higher-order visual cortices appear to process first-order and second-order depth information (such as slant and curvature, respectively) on 3D object structure. Areas involved in processing higher-order depth features of objects in dorsal cortex include hMT+, V3A, V7, and regions along the IPS (Alizadeh et al., 2018; Freud et al., 2016; Georgieva et al., 2009; Janssen et al., 2018; Katsuyama et al., 2011). Many previous studies have found that intraparietal areas are involved in 3D shape processing. For example, there is evidence that the CIP in monkeys, which corresponds to VIPS, V7/IPS0, or V7A in humans (Konen et al., 2013; Orban, 2016), is involved in the representation of 3D curvature defined by disparity (Alizadeh et al., 2018). The CIP region may also be a locus where different types of depth cues are integrated, including disparity and linear perspective (Tsutsui et al., 2001). The CIP area projects to anterior regions of the intraparietal area (AIP) and may belong to a larger object-processing network (Erlikhman et al., 2018). There is also evidence that the AIP region in non-human primates is sensitive to 3D curvature and 3D shape (Joly et al., 2009; Srivastava et al., 2009). The AIP of non-human primates is believed to correspond partially to the human DIPS (Orban, 2016) and Durand et al. (2009) found that some human DIPS regions are sensitive to depth structure. Considering these findings, it is reasonable that accuracy was high in DIPS for all types of classification. In addition, in the two-stream theory proposed by Goodale and Milner (1992), information in the dorsal stream is used for the unconscious control of action, such as the movement of the body guided by visual input. It is possible that the 3D information in the dorsal stream ultimately serves visually guided actions (Cohen and Andersen, 2002; Sakata et al., 1998)). Accordingly, some regions in posterior parietal cortex involved in motor planning respond to depth signals, including AIP for grasping (Snvastava et al., 2009; Theys et al., 2012), the parietal reach region (PRR) (Bhattacharyya et al., 2009), and the lateral intraparietal area (LIP) (Durand et al., 2007). This high classification accuracy independent of depth cue type may be required for rapid visually guided actions. As this information is more

generalized, it does not depend on the specific nature of the depth cue. Therefore, DIPS showed high classification accuracy for both same cue (and element) type and transfer conditions.

3.5.2 Limitations

The main limitation of this study is the potential use of monocular information for concave vs. convex classification. Monocular information may be used for two same cue classification conditions, RDS and lines with disparity, as well as for one transfer classification, RDS/lines with disparity. These monocular cues are explained for the lines with disparity in Figure 3.7 as the RDS condition is similar. Figure 3.7 shows the shapes defined by black–white dotted lines with disparity. For illustration, these images have greater disparity and fewer lines than those used for experiments. Part A shows a stimulus depicting a convex shape, and part B a stimulus depicting a concave shape. For the convex shape, the distance between the left side line and the fixation marker [L1 in (A) of Figure 3.7] is longer than the distance between the right side line and the fixation marker [L2 in (A) of Figure 3.7], i.e., L1 > L2. Alternatively, for the concave shape, the distance between the left side line and the fixation marker [L1’ in (B) of Figure 3.7] is shorter than the length between the right side line and the fixation marker [L2’ in (B) of Figure 3.7], i.e., L1’ < L2’. This difference between convex and concave stimuli may be used by the left eye for the lines with disparity convex vs. concave classification. There is also monocular information accessible to the right eye for classification. For the convex shape, the left side is shorter than the right, i.e., L3 < L4 [(A) in Figure 3.7], whereas for the concave shape, the left side is longer than the right, i.e., L3’ > L4’ [(B) in Figure 3.7]. Similar differences can also be used for the RDS/lines with disparity transfer classification.

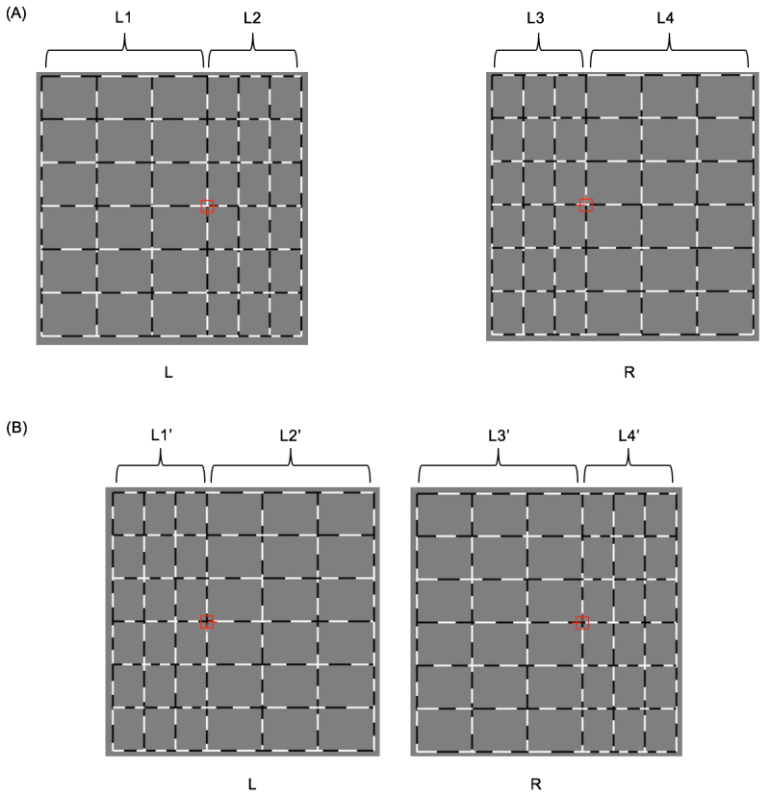

Figure 3.7 Illustration of monocular information for judging convex vs. concave shapes. Shapes defined by black−white dotted lines with disparity are shown. L indicates image projected to the left eye, and R indicates image projected to the right eye. The stimuli shown in this figure are simplified for illustration and are not the real stimuli shown to the participants. (A) Stimuli generating the convex shape. (B) Stimuli generating the concave shape.

Although it is possible that this monocular information was used for classification, we think it is unlikely. First, in the real stimuli for experiments, the length difference between the left side and the right side relative to the fixation marker was only 5.05 mm, which is about 0.2°. Second and more importantly, there was no such monocular information for black−white dotted lines with perspective; therefore, the transfer RDS/lines with perspective and transfer lines with perspective/lines with disparity classifications cannot be made using monocular cues. We conclude that there are common neural activity patterns in the DIPS region for the processing of shapes with different depth cues.

3.6 References of CHAPTER 3

Alizadeh AM, Van Dromme I, Verhoef BE, and Jassen P (2018). Caudal intraparietal sulcus and three-dimensional vision: a combined functional magnetic resonance imaging and single-cell study. Neuroimage 166: 46-59.

Bhattacharyya R, Musallam S, and Andersen RA (2009). Parietal reach region encodes reach depth

using retinal disparity and vergence angle signals. J Neurophysiol.; 102: 805-816.

Chandrasekaran C, Canon V, Dahmen JC, and Kourtzi Z (2006). Neural correlates of disparity-defined shape discrimination in the human brain. J Neurophysiol. 97: 1553-1565.

Clark JJ, and Yuille AL (1990). Data fusion for sensory information processing system. Boston: Kluwer Academic Publishers.

Cohen YE, and Andersen RA (2002). A common reference frame for movement plans in the posteriror parietal cortex. Nature Rev Neurosci.;3: 553-562.

DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser J, Cox R, Miller D, and Neitz J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. Proc Natl Acad Sci USA. 93: 2382-2386.

Dupont P, De Bruyn B, Vandenberghe R, Rosier AM, Michiels J, Marhal G, Mortelmans L, and Orban GA (1997). The kinetic occipital region in human visual cortex. Cereb Cortex 7: 283-292.

Durand JB, Nelissen K, Joly O, Wardak C, Todd JT, Norman JF, Janssen P, Vanduffel W, and Orban GA (2007). Anterior regions of monkey parietal cortex process visual 3D shape. Neuron 55, 493-505.

Durand JB, Peeters R, Norman JF, Todd JT, and Orban GA (2009). Parietal regions processing visual 3D shape extracted from disparity. Neuroimage 15: 1114-1126.

Erlikhman G, Caplovitz GP, Gurariy G, Medina J, and Snow JC (2018). Towards a unified perspective of object shape and motion processing in human dorsal cortex. Consciousness and Cognition 64: 106-120.

Fischl B. FreeSurfer (2012). NeuroImage 62: 774-781.

Freud E, Plaut DC, and Behrmann M. (2016). 'What' is happening in the dorsal visual pathway. Trends Cogn Sci. 20: 773-784.

Freud E, Ganel T, Shelef I, Hammer MD, Avidan G, and Behrmann M (2017). Three-dimensional representations of objects in dorsal cortex are dissociable from those in ventral cortex. Cerebral Cortex 27: 422−434.

Georgieva S, Peeters R, Kolster H, Todd JT, and Orban GA (2009). The processing of three-dimensional shape from disparity in the human brain. J Neurosci. 29: 727-742.

Goncalves NR, Ban H, Sánchez-Panchuelo RM, Francis ST, Schluppeck D, and Welchman AE (2015). 7 tesla FMRI reveals systematic functional organization for binocular disparity in dorsal

visual cortex. J Neurosci. 35: 3056-3072.

Goodale MA, and Milner AD (1992). Separate visual pathways for perception and action. Trends in Neurosciences 15: 20-25.

Hubel DH, and Wiesel TN (1959). Receptive fields of single neurons in the cat's striate cortex. J Physiol. 148: 574-591.

Hubel DH, and Wiesel TN (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol. 160: 106-154.

Janssen P, Vogels R, and Orban GA (1999). Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. Proc Natl Acad Sci U S A. 96: 8217–8222.

Janssen P, Verhoef BE, and Premereur E (2018). Functional interactions between the macaque dorsal and ventral visual pathways during three-dimensional object vision. Cortex 98: 218-227.

Joly O, Vanduffel W, and Orban GA (2009). The monkey ventral premotor cortex processed 3D shape from disparity. Neuroimage 47: 262-272.

Katsuyama N, Usui N, Nose I, and Taira M (2011). Perception of object motion in three-dimensional space induced by cast shadows. Neuroimage 54: 485-494.

Konen CS, Mruczek RE, Montoya JL, and Kastner S (2013). Functional organization of human posterior parietal cortex: grasping- and reaching-related activations relative to topographically organized cortex. J Neurophysiol. 109: 2897-2908.

Kourtzi Z, and Kanawisher N (2000). Cortical regions involved in perceiving object shape. J Neurosci. 20: 3310-3318.

Kourtzi Z, and Kanwisher N (2001). Representation of perceived object shape by the human lateral occipital complex. Science 293: 1506-1509.

Logothetis NK, Pauls J, Augath M, Trinath T, and Oeltermann A (2001). Neurophysiological investigation of the basis of the fMRI signal. Nature 412: 1507.

Maloney LT, and Landy MS (1989). A statistical framework for robust fusion of depth information. Visual Communications and Image Processing IV 1199: 1154-1163.

Mishkin M, and Ungerleider LG (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Behavioural Brain Res. 6: 57-77.

Nelissen K, Joly O, Durand JB, Todd JT, Vanduffel W, and Orban GA (2009). The extraction of depth structure from shading and texture in the macaque brain. PLoS One 4: e8306.

Orban GA (2016). Functional definitions of parietal areas in human and non-human primates. Proceedings of the Royal Society B: Biological Sciences 283: 20160118.

Preston TJ, Li S, Kourtzi Z, and Welchman AE (2008). Multivoxel pattern selectivity for perceptually relevant binocular disparities in the human brain. J Neurosci. 28: 11315-11327.

Sakata H, Taira M, Kusunoki M, Murata A, Tanaka Y, and Tsutsui K. (1998). Neural coding of 3D features of objects for hand action in the parietal cortex of the monkey. Philos Trans R Soc Lond B Biol Sci. 353, 1363-1373.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, and Tooltell RB (1995). Borders of multiple visual areas in human revealed by functional magnetic resonance imaging. Science 268: 889-893.

Srivastava S, Orban GA, De Mazière PA, and Janssen P (2009). A distinct representation of three-dimensional shape in macaque anterior intraparietal area: fast, metric, and coarse. J Neurosci. 29: 10613-10626.

Taira M, Nose I, Inoue K, and Tsutsui K (2001). Cortical areas related to attention to 3D surface structures based on shading: an fMRI study. NeuroImage 14: 959−966.

Theys T, Srivastava S, Van Loon J, Goffin J, and Janssen P (2012). Selectivity for three-dimensional contours and surfaces in the anterior intraparietala area. J Neurophysiol. 107: 995-1008.

Todd JT. The visual perception of 3D shape (2004). Trends in Cogn Sci. 8: 115-21.

Tootell RBH, Hadjikhani N, Hall EK, Marrett S, Vanduffel W, Vaughan JT, and Dale AM (1998). The retinotopy of visual spatial attention. Neuron 21: 1409-1422.

Tsutsui K, Jiang M, Yara K, Sakata H, and Taira M (2001). Integration of perspective and disparity cues in surface-orientation-selective neurons of area CIP. J Neurophysiol. 86: 2856-2867.

Tyler CW, Likova LT, Chen CC, Kontsevich, LL, Schira MM, and Wade AR (2005). Extended concepts of occipital retinotopy. Curr Med Imaging 1: 319-329.

Vanduffel W, Fize D, Peuskens H, Denys K, Sunaert S, Todd JT, and Orban GA (2002). Extracting 3D from motion: differences in human and monkey intraparietal cortex. Science 298: 413-415.

Warnking J, Dojat M, Guérin-Dugué A, Delon-Martin C, Olympiesff, S., Richard, N., Chéhikian, A., and Segebarth C (2002). fMRI retinotopic mapping—step by step. NeuroImage 17: 1665-1683.

Welchman AE, Deubelius A, Conrad V, Bülthoff HH, and Kourtzi Z (2005). 3D shape perception from combined depth cues in human visual cortex. Nature Neurosci. 8: 820-827.

Woods RP, Grafton ST, Holmes CJ, Cherry SR, and Mazziotta JC (1998). Automated image registration: I. General methods and intrasubject, intramodality validation. J Comput Assist Tomogr. 22: 139-152.

Zeki S. Functional specialization in the visual cortex of the rhesus monkey. Nature 1978;274: 423-28.

Zeki S, Watson JD, Lueck CJ, Friston KJ, Kennard C, and Frackowiak RS (1991). A direct demonstration of functional specialization in human visual cortex. J Neurosci. 11: 641-649.

Zeki S, Perry RJ, and Bartels A (2003). The processing of kinetic contours in the brain. Cereb Cortex 13: 189-202.

CHAPTER 4

GENERAL DISCUSSION

In our research, we investigated to main topics: the first topic is the representation of convex–concave shape and horizontal–vertical orientation of stereoscopic curved surface; the second topic is the representation of convex–concave 3D shape consisted of two slanted planes which defined by disparity and perspective respectively. How can we understand these results? What are the relationship between our research and previous researches of others'? In this chapter, we will discuss these questions.

4.1 3D information processing among cortices

The processing of shape and orientation of 3D shape is vital for the survive of human. One of the most important theories about the visual information processing is "two-stream" theory. According to the theory, visual information is progressively through the retina, thalamus, and primary visual cortex (V1) in occipital cortex and then divided into two anatomically and functionally separate streams, a dorsal stream from V1 to the parietal cortex and a ventral stream from V1 to the temporal cortex. Each stream processes visual information in a hierarchical manner, with each cortical region processing information based on the results (output) of the lower regions. Mishkin and Ungerleider (1982) first proposed this theory based on lesion research in non-human primates. The ventral stream has been termed the "what" stream as it processes information related to an object's identity (e.g., shape, size, color, and texture) whereas the dorsal steam has been termed the "where" stream as it processes visual information related to object location, movement, and spatial relationships. Goodale and Milner (1992) proposed a revised version of two-stream theory. Rather than viewing both streams as contributing to conscious visual awareness, they claimed that only the ventral stream contributes to conscious vision (known as the "perception" stream) while information in the dorsal stream is used for the unconscious control of action, such as the movement of the body guided by visual input (and so is known as the "action" stream). From this point of view, 3D shape processing can be occurred in both of the streams, with information in the ventral stream serve for shape

perception while information in the dorsal stream serve for vision-guided action.

During the processing of 3D information along the streams, it is thought that early visual areas mainly process simple attributes of objects, such as orientation, motion direction, and speed; whereas the visual attributes processed by higher visual areas are much more complex, culminating in recognition of objects and complex shapes. Therefore, for "same-type stimuli convex–concave classification" in our first topic, and "classification of convex vs. concave 3D images with the same depth cue" in our second topic, classification accuracy was higher than the baseline of statistical significance in both some of early visual areas and higher visual areas. And we hypnotized that for early visual areas, the high accuracy is based on selectivity for simple attributes of stimuli, while for the high areas, the high accuracy is based on selectivity for the overall information about convex–concave shapes.

(1) Processing shape from disparity

In our study of first topic, we performed three types of convex–concave classification: "same-type stimuli convex–concave classification," "transfer convex–concave classification of surfaces at the same depth position," and "transfer convex–concave classification of surfaces at different depth position." We found that in V3A, classification accuracy is higher than the baseline of statistical significance. This results indicate that V3A play is an important role in the stereoscopic information processing along the visual streams. This is consistent with previous related studies:

First, V3A is selective for disparity. In a study of Preston et al. (2008), they adopted MVPA method on fMRI data, and found that V3A is selective for depth position of plane defined by disparity. The selectivity for disparity is one of possible reason for high accuracy in "same-type stimuli convex–concave classification."

Second, Coarse stereopsis is processed in the dorsal steam and fine stereopsis in processed in the ventral stream (Neri et al., 2004; Uka & DeAngelis, 2006; Roe et al., 2007; Schiller et al., 1990). In addition, V3A is on the dorsal visual stream, and this visual stream is related to coarse stereopsis. It is likely that V3A was involved in the representation of the shape in a more abstract (coarse) manner, which did not depend on the orientation of the surface and depth position.

Third, it is possible that the coarse information in the dorsal stream is needed by the ventral stream for perceive fine stereopsis and information exchange is required between dorsal and ventral stream. Many evidence showed that there is information exchange between the dorsal and ventral streams and V3A is one of the areas for information exchange. In the anatomical study of Takemura et al. (2016), they identified a major white pathway, the vertical occipital fasciculus (VOF), connecting

maps within the dorsal and ventral visual cortex. They used a model-based method to assess the statistical evidence supporting several aspects of the VOF wiring pattern. They claimed that there is strong evidence supporting the hypothesis that dorsal and ventral visual maps communicate through the VOF, and the cortical projection zones of the VOF suggest that human ventral (hV4/VO-1) and dorsal (V3A, KO) maps exchange substantial information. Also, for functional study, Sim et al. (2015) investigated the functional contribution of the visuo-motor system to visual object recognition using fMRI and ERP during action priming. Effective connectivity analyses suggested functional influences of parietal areas (dorsal stream) on anterior temporal area (ventral stream). In this consideration, V3A may be a locus to transfer coarse stereopsis information to ventral stream, and it therefore showed high accuracy in all types of convex–concave classification in the first topic of our study.

(2) Processing shape from different cues (disparity or perspective)

For shape from perspective, similar to disparity, early visual areas are thought process elements of the shape from perspective (such as orientation of lines), higher areas process more complex factors of shape. This is also supported by our results in our study of the second topic: same cue type classification showed high accuracy in both early and high visual areas, whereas transfer convex vs. concave classification on stimuli of different type types of cues showed high accuracy mainly in DIPS. As the 3D shape information from different cues processed along the visual pathways, these depth cues concurrently contribute to 3D shape perception, and many models have been proposed to explain how different depth cues contribute to 3D shape processing at the neural level. Maloney and Landy (1989) proposed a simple statistical framework for combining depth estimates from consistent depth cues in which information from different cues is processed independently by different modules and then fused into a single depth estimate about each point of the scene. They also assumed that the assigned weights of different estimates are variable and that the combination is linear (additive). In contrast, Clark and Yuille (1990) distinguished between “strong fusion” and “weak fusion”. In strong fusion, information from different cues interacts and is processed cooperatively to yield a single depth estimate, whereas in weak fusion information from different cues is independent and the final estimate is obtained by combining the individual information. Although there is no consistent idea on how depth cues work together, it commonly agrees that a final depth estimate of each point of shape will generated at a certain stage, and 3D information is less dependent on the concrete depth cues. Specifically, in the dorsal stream, visual information processing follows an overarching posterior-to-anterior visual-to-motor gradient, with more posterior areas dedicate predominantly to analysis of visual properties, while anterior areas are increasingly concerned with computations related to

potential motor interactions with the stimulus, such as encoding the object's absolute location in egocentric space, reaching trajectory, and hand pre-shaping (Fabbri et al., 2016; Stark et al., 2008). In the second topic of our study, classification accuracy was higher than the baseline of statistical significance in transfer convex vs. concave 3D images generated with different depth cues in DIPS which indicate common neural activity pattern is involved in the processing 3D information from disparity depth cue and perspective depth cue. It is possible this common activity pattern is related to the potential motor interactions with the shapes, and this information is not related to the concrete depth cue type.

4.2 References for CHAPTER 4

Clark JJ, and Yuille AL (1990). Data fusion for sensory information processing system. Boston: Kluwer Academic Publishers.

Fabbri S, Stubbs KM, Cusack R and Culham JC (2016). Disentangling representations of object and grasp properties in the human brain. J Neurosci. 36, 7648-7662.

Mishkin M, and Ungerleider LG (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Behav. Brain Res. 6, 57–77.

Goodale MA, and Milner AD (1992). Separate visual pathways for perception and action. Trends Neurosci. 15, 20-25.

Maloney LT, and Landy MS (1989). A statistical framework for robust fusion of depth information. Visual Communications and Image Processing IV 1199: 1154-1163.

Neri P, Bridge H, and Heeger DJ (2004). Stereoscopic processing of absolute and relative disparity in human visual cortex. J. Neurophysiol. 92, 1880-1891.

Preston TJ, Li S, Kourtzi Z, and Welchman AE (2008). Multivoxel pattern selectivity for perceptual relevant binocular disparities in the human brain. J. Neurosci. 28, 11315-11327.

Takemura H, Rokem A, Winawer J, Yeatman JD, Wandell BA, and Pestilli F (2016). A major human white matter pathway between dorsal and ventral visual cortex. Cereb. Cortex 26, 2205–2214.

Roe AW, Parker AJ, Borm RT, and DeAngelis GC (2007). Disparity channels in early vision. J. Neurosci. 27, 11820-11831.

Schiller PH, Logothetis NK, and Charles ES (1990). Role of the color-opponent and broad-band channels in vision. Visual Neurosci. 29, 321-345.

Sim EJ, Helbig HB, Graf M, and Kiefer M (2015). When action observation facilitates visual perception: activation in visuo-motor areas contributes to object recognition. Cereb Cortex 25, 2907-2918.

Stark A, and Zohary E (2008). Parietal mapping of visuomotor transformations during human tool grasping. Cereb Cortex 18, 2358-2368.

Uka T, and DeAngelis GC (2006). Linking neural representation to function in stereoscopic depth perception: roles of the middle temporal area in coarse versus fine disparity discrimination. J. Neurosci. 26, 6791-6802.

Appendix

A. Define the ROIs

The ROIs include retinotopic areas V1, V2, V3d, V3v, and V3A, higher dorsal areas KO, V7, and hMT+, higher ventral areas: LOC, IPS areas: DIPS, POIPS, and VIPS. All the stimuli we used were generated using Psychotoolbox 3 in MATLAB (The MathWorks).

A.1 Retinotopic Mapping

V1, V2, V3d, V3v, and V3A is localized by a rotating wedge (Figure A.1-A) and an expanding ring (Figure A.1-B). The rotating wedge stimulus is used to delineate the retinotopic areas. The angle of the wedge is 45°, and the angular velocity of rotating is 11.25 deg/s. Each cycle lasts 32 s, and for each run includes 8 cycles. The expanding ring stimulus, the ring expands from the center to the periphery, it is used to determine the eccentricity axis from the center to periphery. Each cycle lasts 32 s, and for each run includes 8 cycles. There is a fixation point in the center of both the rotating wedge stimulus and the expanding ring stimulus which is used to help participants to maintain vergence and avoid eye movement.

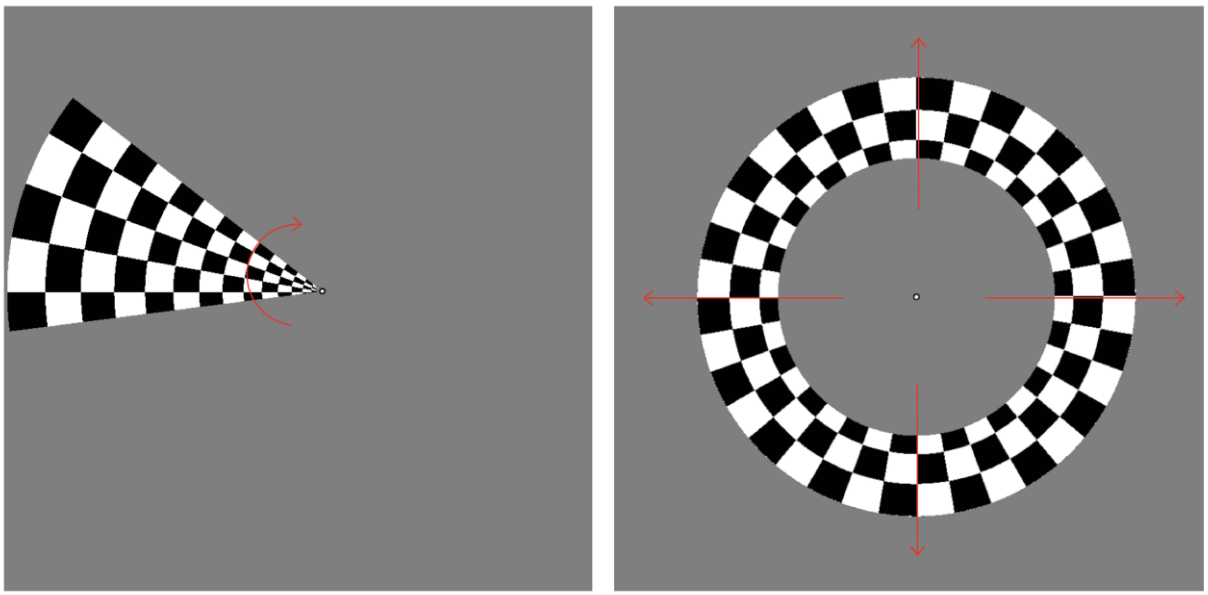

Figure A.1 Illustration of stimuli for retinotopic mapping. (A) Stimulus of single a wedge. (B) Stimulus of a ring.

A.2 Localizing higher dorsal areas

The higher dorsal areas which we investigated in our experiments include hMT+ and KO. Figures A.2 shows the stimuli to localize hMT+ area. The localizer of hMT+ is similar to the one used by Huk et al. (2002). It consists of two types of stimuli: randomly generated static white dots within a circle area (Figure A.2-A); randomly generated white dots which are coherently moving outward or inward (Figure A.2-B). The hMT+ is defined as an area which shows more activated to stimulus the moving dots compared to the static dots. In the center of the screen, there is a small white square which works as fixation marker to help participants to keep the vergence and avoid eye movement.

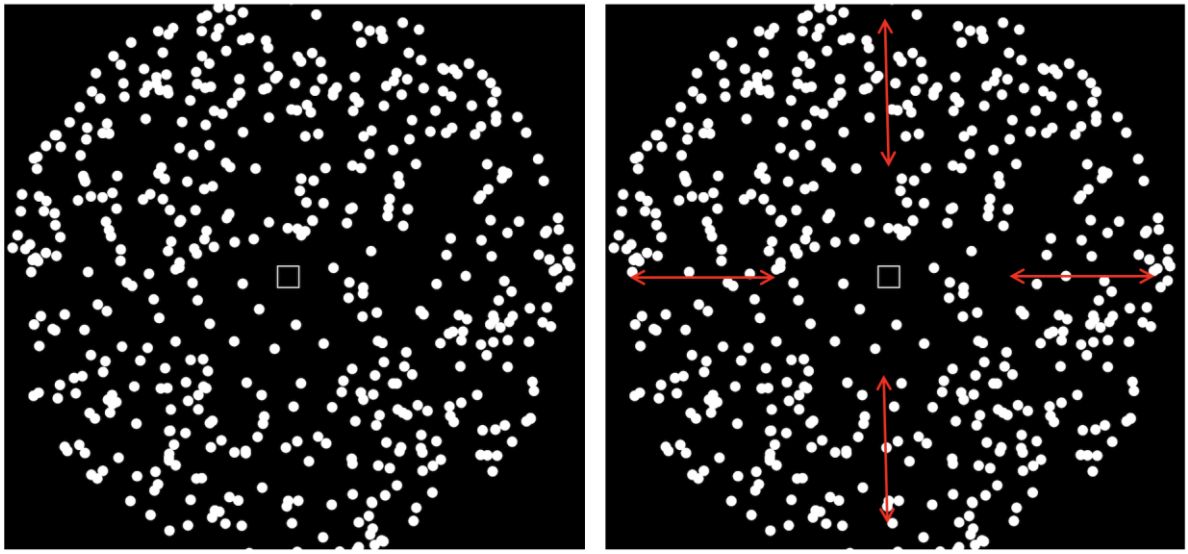

Figure A.2 Illustration of hMT+ localizer.

Figure A.3 shows the localizer of KO. It is similar to the studies of Dupont et al. (1997) and Van Oostende et al. (1997). It contains two types of stimuli: motion-defined contours of a field of black and white dots (Figure A.3-A); transparent motion of a field of black and white dots (Figure A.3-B). KO is defined by a continues areas which shows more activated to stimulus of motion-defined contours compare to transparent motion. There is a black dot in the center of the screen which works as a fixation point in order to keep the vergence of the eyes and avoid eye movement.

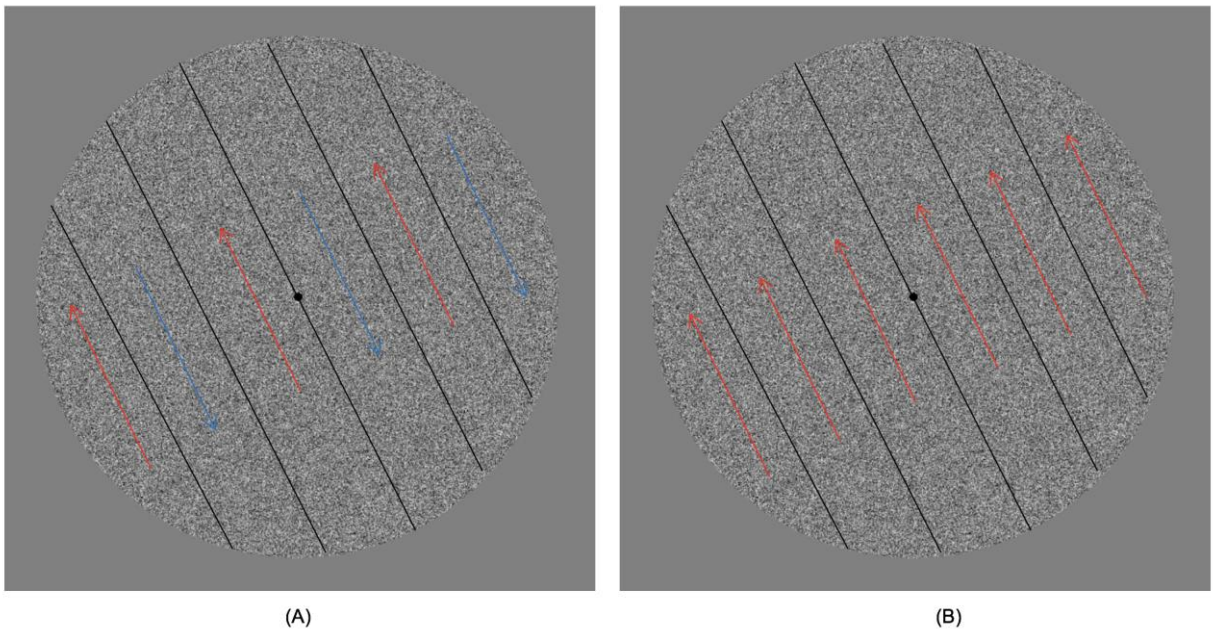

Figure A.3 Illustration of the KO localizer.

A.3 Localizing higher ventral areas

We localized higher ventral area LOC. The localizer we used is similar with stimuli used in the study of Kourtzi and Kanwisher (2000). The localizer of LOC contains two types of stimuli: the intact object (Figure A.4-A, B, C, D) and the scrambled version of the object (Figure A.4-A', B', C', D'). The intact object is depicted by either grayscale image (Figure A.4-A, B) or line drawing (Figure A.4-C, D) image. The objects of images can be familiar (Figure A.4-A, C) or novel (Figure A.4-B, D) to the participants. The scramble version of object is generated as follows: first, divide the intact images into a 20 × 20 square grid; second, shift the positions of squares randomly.

A block design was used to show these stimuli. Each run consisted of sixteen stimulus epochs with fixation epochs interleaved. Each stimulus epochs lasted for 16 s. Twenty different images of the same type (intact or scramble) were shown in each stimulus epochs. Each image was shown for 0.25 s with a black interval of 0.55 s after it. During the scans, the participants were required to performed a fixation task and judged whether they familiar with each intact object or not. The LOC area is defined as a set of continuous voxels in the lateral occipito-temporal cortex that showed stronger activation to intact versus scrambled images of objects.

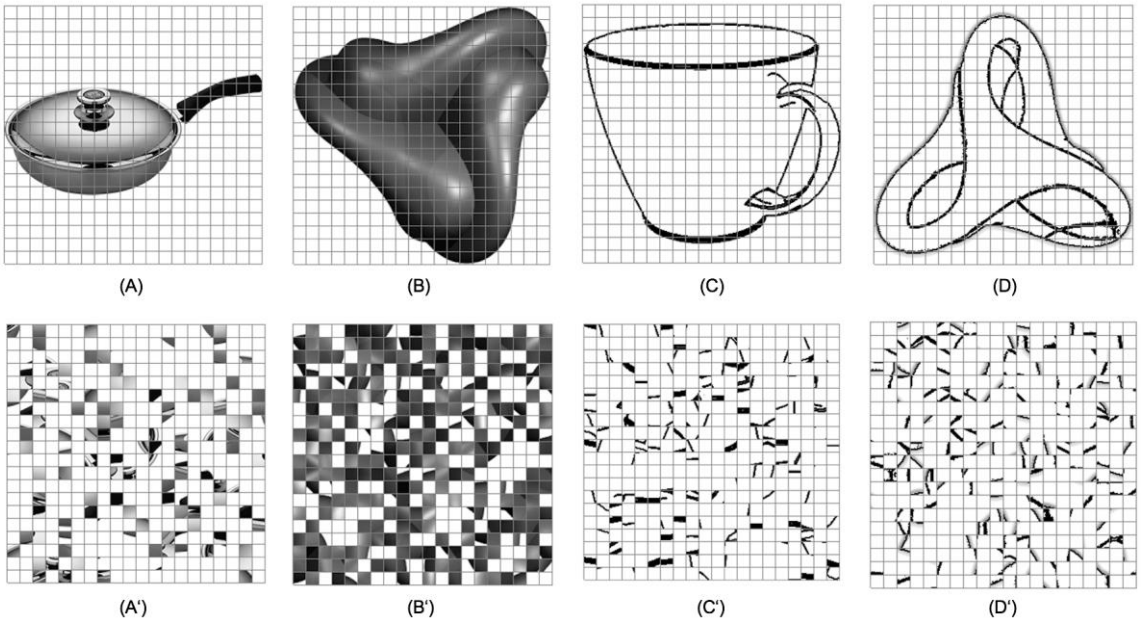

Figure A.4 Illustration of the LOC localizer.

A.4 Localizing IPS areas

We adopted the similar stimuli which were used in the studies of Orban et al. (1999) and Vanduffel et al. (2002). The stimuli consist of nine connected lines. There are two conditions for our stimuli: (1) the stimuli move in the fixation plane, including translating in along the horizontal and vertical axes, and rotating along depth axis; (2) in addition to move in the fixation plain, the stimuli also rotate in depth for 3D shape from motion, including rotating along horizontal and vertical axes. During the scan, the participants were required to observe the fixation marker. The IPS areas are identified by contracting activity to condition 2 to activity to condition 1.

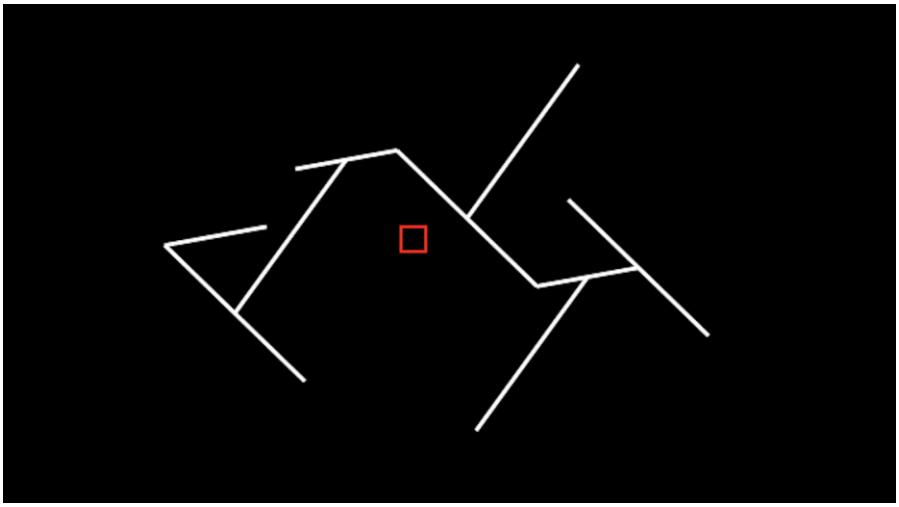

Figure A.5 Illustration of the stimulus to localize IPS.

B. Abbreviations

AIP anterior intraparietal area

BOLD blood oxygenation level-dependent

CIP caudal intraparietal area

DIPS dorsal intraparietal sulcus

EPI echo-planar imaging

fMRI functional magnetic resonance imaging

GM gray matter

IPS intraparietal sulcus

KO kinetic occipital area

LGN lateral geniculate nuclei

LOC lateral occipital complex

MT middle temporal

MVPA Multi-voxel pattern analysis

hMT+ human middle temporal complex

POIPS parieto-occipital intraparietal sulcus

ROI region of interest

RDS rondom dot stereogram

SVM support vector machine

TE temporal sulcus

VIPS ventral intraparietal sulcus

WM white matter

C. References for Appendix

Dupont P, De Bruyn B, Vandenberghe R, Rosier AM, Michiels J, Marchal G, Mortelmans L, Orban GA (1997). The kinetic occipital region in human visual cortex. Cereb Cortex. 7: 283-292.

Huk AC, Dougherty RF, and Heeger DJ (2002). Retinotopy and functional subdivision of human areas MT and MST. J Neurosci. 22: 7195-7205.

Kourtzi Z., and Kanawisher N (2000). Cortical regions involved in perceiving object shape. J. Neurosci. 20, 3310-3318.

Orban GA, Sunaert S, Todd JT, Van Hecke P, and Marchal G (1999). Human cortical regions involved in extracting depth from motion. Neuron 24: 929-940.

Vanduffel W, Fize D, Peuskens H, Denys K, Sunaert S, Todd JT, and Orban GA. (2002). Extracting 3D from motion: differences in human and monkey intraparietal cortex. Science 298, 413-415.

Van Oostende S, Sunaert S, Van Hecke P, Marchal G, and Orban GA. (2002). The kinetic occipital (KO) region in man: an fMRI ftudy. Cereb Cortex. 7, 690-701.

LIST OF PUBLICATIONS

A. Journal Paper

Li Z, and Shigemasu H. Generalized representation of stereoscopic surface shape and orientation in the human visual cortex (2019). Front Hum Neurosci. 13: 283. (JCR Q2)

B. Conferences

Li Z, and Shigemasu H. Decoding of 3D shape perception from human brain activity. International Workshop on Human-Engaged Computing, Kochi, Japan. November, 10-11, 2017. (Poster)

Li Z, and Shigemasu H. Generalized representation of stereoscopic surface in V3A. Annual meeting of Vision Sciences Society, St Pete Beach, Florida, USA. May, 18-23, 2018. (Poster)

Li Z, and Shigemasu H. Generalized representation of shapes from different cues. International Workshop on Human-Engaged Computing, Kochi, Japan. January, 12, 2019. (Poster, **best poster award**)

Threethipthikoon T, **Li Z**, and Shigemasu H. Generalized representation of 3D object related to action in visual cortex: an fMRI study. International Workshop on Human-Engaged Computing, Kochi, Japan. January, 12, 2019. (Poster)

Li Z, and Shigemasu H. Generalized representation of shapes from different cues in parts of IPS areas. Annual meeting of Vision Sciences Society, St. Pete Beach, Florida, USA. May, 17-22, 2019. (Poster)

Li Z, and Shigemasu H. Common cortical representation of convex–concave shapes from different depth cues. The 15th Asia-Pacific Conference on Vision, Osaka, Japan. July, 29–August, 1, 2019. (Poster)

Threethipthikoon T, **Li Z**, and Shigemasu H. Generalized representation of 3D object orientation in human visual cortex. The 15th Asia Pacific Conference on Vision, Osaka, Japan. July, 29–August, 1, 2019. (Poster)

ACKNOWLEDGMENT

Three years is a long time, but also a short time. As a doctoral student during the past three years, my life was full of hardness and joys. On the time of finishing the dissertation, I would like to express my sincere gratitude to a lot of people who have helped me, supported me and encouraged me, without whom I could not finish my dissertation.

First and foremost, I would like to express my sincerest thanks to my supervisor Assoc. Prof. Hiroaki SHIGEMASU who accepted me as a member of his laboratory and guided me into the field of 3D vision and brain information processing. He is a kind and patient person with enthusiasm and immense knowledge. He gave me countless advice on my research, from designing experiments to writing papers.

Besides my supervisor, I would like to express my deep thank to co-supervisors: Prof. Keizo SHINOMORI and Assoc. Prof. Hiroshi KADOTA for the enlightening discussion and time spent on my research. My sincere thanks also go to my dissertation assessment committee members: Prof. Kiyoshi NAKAHARA and Assoc. Prof. Yukinobu HOSHINO who spent their time on reviewing my dissertation and gave significant comments on it.

Thanks to all participants of my fMRI experiments. During these experiments, they spent a lot of time. Without their cooperation, my work would not be finished.

Special thanks also go to my friends at Kochi University of Technology, including Benchun CAO, Fangyuan LIAO, Jingxin LIU, Lilin XIE, and Zongrong XIN (alphabetical order). We helped with others in daily life and this made me feel comfortable.

Last but not least, I would like to show my deep gratitude to my family members: including my parents, father-in-law, mother-in-law, my wife, and my daughter. My parents always encouraging me; my father-in-law, mother-in-law and my wife helped me take care of my daughter, it was very hard for them; my daughter is lovely and makes happy every day. Because of their supports, I could concentrate on my research.

Zhen LI

2019/09