NAIST-IS-MT1051207

Master's Thesis

Neural Decoding of Visual Object Position from Human fMRI data

Sukhanov Paul

August 16, 2012

Department of Information Systems
Graduate School of Information Science
Nara Institute of Science and Technology

A Master's Thesis
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
MASTER of ENGINEERING

Sukhanov Paul

Thesis Committee:

Professor Kazushi Ikeda	(Supervisor)
Professor Mitsuo Kawato	(Co-supervisor)
Professor Yukiyasu Kamitani	(Co-supervisor)
Associate Professor Kenichi Matsumoto	(Co-Supervisor)

Neural Decoding of Visual Object Position from Human fMRI data*

Sukhanov Paul

Abstract

Visual information in the early primate visual system is known to be represented in a retinotopic fashion, so that adjoining points in the visual field are represented by neural activities at adjoining points on the cortical sheet. Although a rough form of retinotopy persists into higher visual cortex, the way in which spatial information is encoded in higher areas has not been thoroughly explored. In particular, the way in which the degree of retinotopy and the sizes of receptive fields in higher visual cortex constrain the accuracy of spatial encoding is not well understood. In the current study we aimed to investigate these questions by decoding the spatial position of 3-dimensional moving objects from fMRI activity recorded during observation of the objects by human subjects. This was achieved by training sparse linear regression models using neural activity from different areas of visual cortex and observing the accuracy of predicted object trajectories for a test set of fMRI data. Notably, accurate predictions were obtained even from higher level visual cortex. These accuracies are then viewed in terms of the sizes of receptive fields known to exist in each area, and a computational model is used to demonstrate how wide receptive fields are compatible with accurate spatial encoding.

Keywords:

Neural Decoding, Retinotopy, Spatial Information, Receptive Field, fMRI

*Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT1051207, August 16, 2012.

Contents

List of Figures

List of Tables

1. Introduction

1.1 Overview

When humans observe a visual scene, it is quickly apparent to us the locations of the objects and other visual features that we are seeing. However, the process by which the visual information is transformed into spatial coordinates is complicated and the question of how the human visual cortex processes and represents this position information is a broad topic of study in neuroscience. It is known that human visual cortex consists of multiple areas, each of which contains a "map" of the same visual space [1]. However, differences in the scale of the visual information being represented in each map can have implications for the accuracy with which position information is represented. Assessing the accuracy with which each map encodes the positions of its respective visual features in the face of increasing complexity and changes in scale is the main topic of this thesis.

1.2 Human Visual System

1.2.1 Visual Processing Hierarchy

The visual world that we perceive seems naturally to us to be composed of many separate objects together forming a coherent scene. The partitioning of the array of light impinging on the retina into discrete objects and groups of objects is however a complicated process and is separated into several stages of successive processing (see Fig. 1). At each stage of the hierarchy, the activity of neurons represents different types of information about the same visual scene at different scales. For example, the level of activity of photoreceptors in the retina corresponds directly to the luminance of light (of specific wavelengths) at each point across the eye, while activity in the next visual area, LGN, is generated from incorporating several outputs of the retina and represents local light contrast (small dots) [2].

Likewise, neurons located further along the hierarchy accept input from several neurons at the previous (and same) stage, and can represent more complex features, such as oriented edges in area V1[3], specific shapes in area V4, or even particular categories of objects in ventral temporal cortex (VTC). For example,

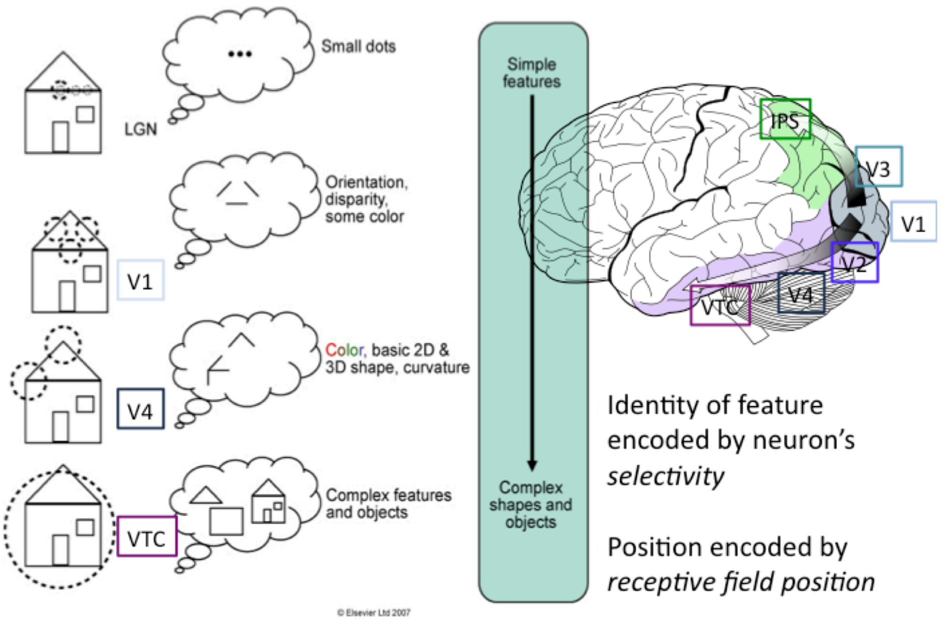
Small dots
LGN
Orientation, disparity, some color
V1
Color, basic 2D & 3D shape, curvature
V4
Complex features and objects
VTC
© Elsevier Ltd 2007
Simple features
Complex shapes and objects
IPS
V3
V1
V2
VTC
V4
Identity of feature encoded by neuron's *selectivity*
Position encoded by *receptive field position*

Figure 1. Visual Processing Hierarchy.

Adapted from [22]. As one progresses through successive visual areas in the human brain, the information represented becomes more complex and receptive fields become larger.The diagram shows sample receptive fields of neurons from the labeled visual areas on the left, and the type of visual feature encoded by the neurons in the center. The corresponding region of the brain is highlighted on the far right.

there are neurons in an advanced stage of the hierarchy whose activity is chiefly modulated by visual stimuli that resemble a face (Fusiform Face Area (FFA), [4]), house (Parahippocampal Place Area (PPA) [5], or other specific category of object (LOC,[6]). The region of space that a visual feature must be present in so that the neuron will react is called the neuron's ”receptive field” (RF). Both the complexity of the visual features and the sizes of the RFs increase as one moves further up the hierarchy.

1.2.2 Position Encoding and Retinotopy

Since the time of the lesion studies of Holmes [7, 20], it has been known that damage to particular areas of the brain results in loss of vision in specific parts of the visual field. This occurs because there is a topological mapping from the input at the retina to the position of RFs of neurons in many areas of visual cortex. This organization in which adjacent neurons possess RFs that are placed in adjacent points of the visual field is known as a ”retinotopic” organization, or sometimes ”visual field map” [1]. The activity of neurons in specific locations of each of the visual areas explained above represents the presence of some visual feature in the corresponding region of space. However, the type of feature encoded in each map is different, and the spatial scale of the feature is also different.

1.2.3 Ventral and Dorsal Streams

When the visual feature being encoded is relatively simple (such as an oriented line segment), the above-mentioned retinotopic encoding scheme is a very straightforward way to simultaneously encode the position and identity of the visual stimulus. However, when the visual feature is more complicated, such as an entire object, the number of neurons required at each position of the visual field would multiply to a prohibitively high number (one for every object you could imagine!). One way in which the brain solves this problem is by separating the encoding of identity and position of objects into two different processing ”streams” [8, 9]. A ”ventral” or ”what” stream is composed of a series of areas associated with creating an invariant representation of the identity of a given visual feature, so that neurons in those areas will respond selectively to particular categories of objects regardless of their position in the visual field, closeness or

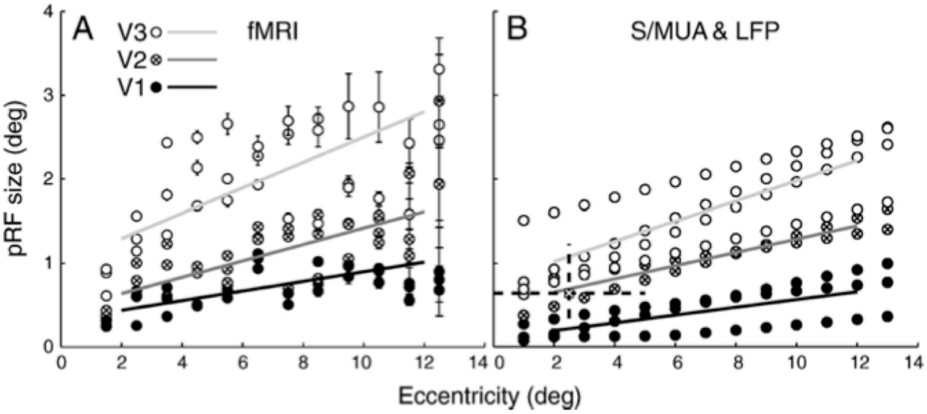

Figure 2. Receptive field sizes by visual area.

Adapted from [11]. population receptive field (pRF) modeling was conducted to estimate the pRF size for different visual areas. The sizes of estimated pRFs by eccentricity is shown on the left, and the equivalent estimate based on single or multi-unit and LFP recordings is shown on the right. The markers and lines are labeled with the corresponding visual areas.

farness from the observer, or angle of presentation. An important caveat however is that even though most neurons have receptive fields large enough to be make them partially invariant to the position of the stimulus, a complete invariance is not achieved [10]. A second "dorsal" or "where" stream has been strongly implicated in the processing of object position, specifically for processing the targets of human interaction.

1.3 Problem Statement

It is known that the average receptive field sizes of neurons increase as one ascends the visual processing hierarchy either along the ventral stream or the dorsal stream ([11], also see Fig. 2). It is also often assumed that with the increase in RF size, the amount of spatial information decreases because large RFs are the hallmark of neuronal position invariance to stimuli (same neural response regardless of position of stimulus). While previous studies have quantified the level of invariance of individual neurons to stimulus position [1] , there has not been an exploration of the level of spatial or position information in an area as a whole.

Recently the advent of machine-learning based neural decoding has allowed for the readout of observed stimulus characteristics directly from multivoxel brain patterns, making it possible for the first time to explore the question experimentally in humans. The goal of the current study is to expand our knowledge of how accurately visual stimulus position is encoded in different areas of visual cortex by decoding visual position information directly from human brain activity.

This was accomplished by first collecting brain activity data using fMRI for two human subjects as they observed a moving sphere object traverse a screen. Then using a statistical method (see section 2.6 for details), a relationship between the brain activity pattern in different visual areas at a particular time and the x and y coordinates of the ball at that same time was learned. A separate set of test data was used to evaluate the performance of the model trained using data from each visual area. In this way, a measure of the spatial information by area was obtained, and allowed for comparison by the sizes of receptive fields of neurons in each area. We observed a high accuracy of decoding even for areas in which receptive field sizes are known to be quite large (such as the lateral occiptal complex LOC, and intraparietal sulcus IPS) , although the highest accuracy was seen in early visual areas V1 to V3.

To further interpret the high accuracy of position decoding for areas with large receptive fields, a computational model of voxel responses to stimulus position was constructed (section 2.7). We assumed that each voxel could be modeled as having a receptive field with a center on the x-y plane and an activity which decreases exponentially from a peak level (identical for each voxel) as the stimulus moves away from the center of the voxel RF, plus a normally distributed noise term. Uniformly distributed RF centers and randomly placed targets were used, and the simulated voxel responses used to predict the stimulus position with one of two methods (regression or maximum likelihood). The effect of increasing receptive field sizes on the accuracy of the x-y position prediction was examined.

The results of the simulation study showed that in fact there is no dependency of encoding accuracy on the size of the receptive field when the maximum likelihood solution is used (see section 3.2). However we did observe a trend of increasing accuracy with RF size when implementing a linear readout mechanism (linear regression), which agrees with our experimental results within lower visual cortex. These results are theoretically justified by previous research such as the work of Zhang and Sejnowski [12], who showed that for 2-dimensional variables, the amount of information contained in a set of firing neurons does not depend on the width of the tuning curves of the neuorns.

2. Materials and Methods

The general experimental overview is presented in Fig. 3. Two human subjects were shown a display with a moving sphere stimulus while brain activity was recorded with fMRI. At each timepoint, the percent signal change (see section 2.5) of a selection of fMRI voxels was used as input, and the x and y values of the position of the center of the sphere used as target labels, for training a sparse linear regression model.

2.1 Subjects

Two healthy male subjects (right-handed, ages 23 and 24) with normal or corrected-to-normal vision participated in the experiments. Signed, written consent was obtained from each subject and approval obtained from the ATR ethics committee.

2.2 Stimulus and Experimental Protocol

All stimuli were created with Psychtoolbox v.3 for Matlab and the associated openGL for psychtoolbox extension. Stimuli were backprojected onto a display in the fMRI scanner and viewed through a mirror attached to the scannerfs headcoil with a field of view equal to 12.5 × 12.5 degrees. For each subject, four scanning sessions (runs) were conducted. In each run, an initial rest period of 32 s was followed by four blocks of stimulus presentation, each lasting for 240 s and interleaved with 12-s rest periods. During pre-rest and rest periods, a circular fixation point (0.25° diameter) was displayed on the center of the display and subjects were instructed to attend to this point. During stimulus presentation, in addition to the fixation point, a white-and-black checkered sphere of diameter 1.6° was displayed at a flickering rate of 6 Hz (Fig. 3).

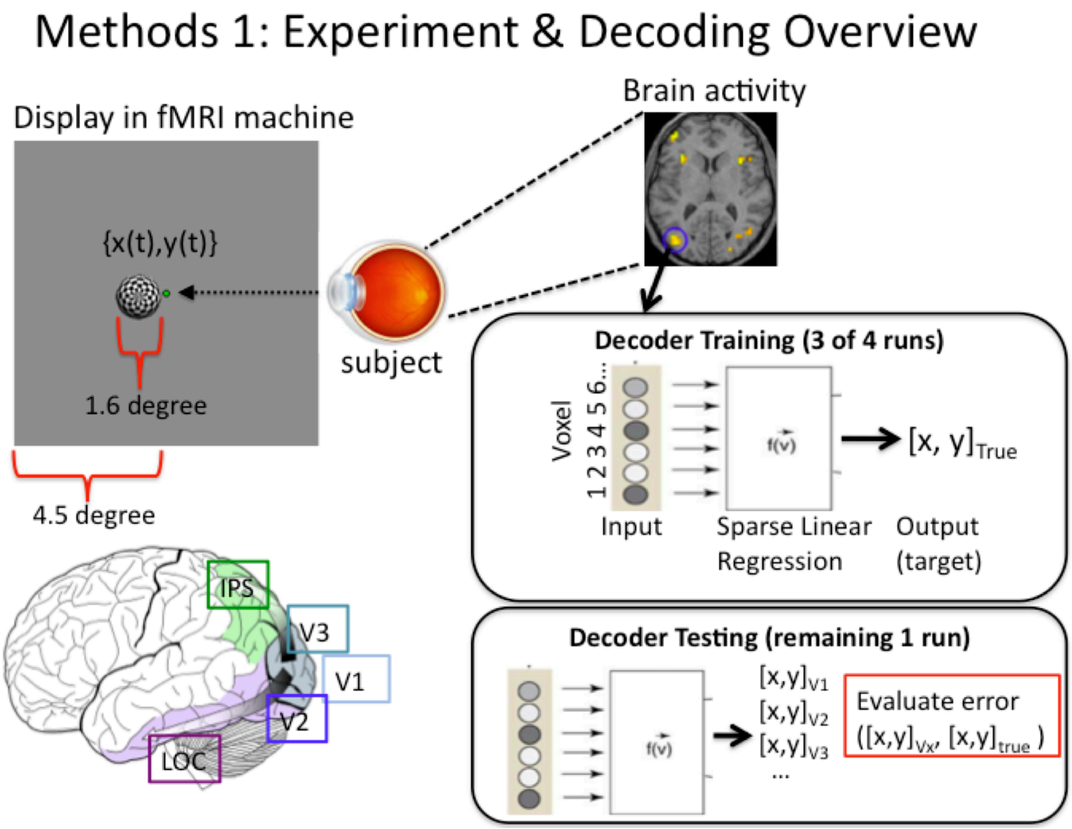

Figure 3. Experimental Overview

The screen shown on the left was viewed by subjects within an fMRI scanner. The notations show the size of the sphere, the size of the maximal radius of movement of the ball and the x and y position of the center of the sphere at a timepoint t (notations not displayed during experiment). The right side of the figure illustrates the decoding procedure, in which a sparse linear regression model is trained to take as input a set of voxels, and output either the x or the y position of the sphere.

Table 1. Stimulus Parameters

parameter	speed(deg/s)	diameter(deg)	radius from center(deg)
min	0.13	0.00	1.31
max	2.20	4.01	2.15
mean	1.00	2.37	1.70

The sphere was programmed to move in a pseudo-random orbit about the central fixation point, by modulating its translation and rotation directions independently using sin and cosine functions: In each frame, first the rate variable t was incremented by $\pi/16$, then the ball was rotated t degrees around the currently facing axis, translated by $R\sin(\omega_1 t)$ on the x-axis (where R is the maximum radius, 4.5° and ω_1 a scaling factor equal to 0.01) and by $R\sin(\omega_2 t)$on the z-axis (relative to the facing direction, ω_2 equal to .005), and finally rotated again by t degrees on the y-z axis relative to facing direction. The result is a smooth but complex and unique trajectory that remains within the bounds of the visual field. The speed and size of the sphere were modulated throughout the course of the 3D orbit (see table 1), which helped keep the subjectfs attention and improved accuracy compared to a 2D, fixed speed orbit (results not shown). During presentation, subjects were instructed to maintain fixation on the central fixation point, but to continue to covertly attend to the position of the sphere.

2.3 Retinotopy and Functional Localizer

Five separate visual areas were identified for use in the decoding analysis. Three lower visual areas (V1, V2, V3) were identified for each hemisphere using a conventional retinotopic mapping experiment[14] and analysis with BrainVoyager QX v.3 [**?**]. In addition, two higher visual areas, Lateral Occipital Complex LOC) and Intraparietal Sulcus (IPS) were defined on the basis of functional localizer experiments previously presented in the literature [4, 15]. For area LOC, a comparison of brain activation in response to scrambled objects vs. intact objects was used to identify voxels ($p < 0.01$; see [4] for more details on experimental procedure) For area IPS, a comparison between brain activity during passive viewing and attentive tracking of objects in a multiple object tracking task was used to define voxels belonging to IPS ($p < 0.01$) [15]. These five areas were selected as "representative" visual areas. Areas V1, V2, and V3 cover early visual cortex, and we assumed the accuracy of LOC would be representative of accuracies from ventral cortex and accuracy of IPS to be representative of accuracies from dorsal cortex.

2.4 fMRI Data Acquisition

fMRI data was recorded on a Siemens Trio 3T scanner using the following scan parameters: TR = 2000 ms, TE = 35 ms, voxel size=$3 \times 3 \times 3$mm, number of slices = 31, slice gap 0 mm.

2.5 fMRI Data Preprocessing

Each volume of functional MRI data was time-slice adjusted, corrected for head movement, and co-registered to the subject's structural MRI image using SPM5. The first four volumes of every run were discarded due to transient magnetic start-up effects. Each voxelfs time course was shifted by 2 samples (4s) to account for the hemodynamic delay, and converted to percent signal change by using the initial 32-s rest period from each fMRI run. Outlier removal and linear detrending was also performed.

2.6 Decoding Analysis

The x and y position of the center of the sphere (measured in visual degrees from the central fixation point) at the middle of every 2-s interval (every volume of brain data) was used to label the fMRI data recorded during that time interval. Decoders for the x and y values were built separately by using the labeled fMRI data from three of the four runs as input to a sparse linear regression algorithm [16] with testing conducted on the fourth run in a leave-one-run-out cross-validation procedure. Because fMRI data is highly multidimensional, a sparse algorithm was required for selecting a subset of relevant features to enable good generalization across data sets. The selected algorithm has already proven to be effective for this purpose and has been tested in the context of extracting voxels for decoding of visual cortical activity. [13]. The method works by applying Baye's rule on equation 1 and maximizing the posterior probability of the weights $\mathbf{w}$ given the data $\mathbf{V}$. An automatic relevance determination (ARD) prior (equation 4) over the weights allows voxels that are unimportant to be automatically removed (weights driven to 0).

$$t_n = \sum_{m=1}^{M} w_m \phi_m(\mathbf{V}) + \varepsilon \tag{1}$$

$$p(\varepsilon|\sigma^2) = N(0, \sigma^2) \tag{2}$$

$$p(\mathbf{t}|v, w, \sigma^2) = \prod_{m=1}^{N} p(t_n|v_n, \mathbf{w}, \sigma^2) \tag{3}$$

$$p(\mathbf{w}|\alpha) = \prod_{m=1}^{M} \left(\frac{\alpha}{2\pi}\right)^{1/2} \exp\left(-\frac{\alpha}{2} w_m^2\right), \tag{4}$$

where $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_N)$ is a data set of N observations of M voxels, $\mathbf{t} = (t_1, \ldots, t_N)$ is a vector of target values, ϕ_m is a fixed basis function, $\mathbf{w}$ is a set of weights, ε is a Gaussian noise term, and α is a (Gaussian) prior distribution over the values of the weights. Voxels from lower (V1, V2, and V3) and higher visual areas (LOC, IPS) were used to build decoders for each visual area independently, which enabled comparison of results across visual areas. The goodness-of-fit between the predicted coordinates and the true coordinates of the sphere throughout the time course were evaluated using the correlation coefficient r between the two trajectories:

$$r = \frac{\sum_{i=1}^{N}(x_i^s - \bar{x^s})(x_i^P - \bar{x_i^P})}{\sqrt{\sum_i^N (x_i^s - \bar{x^s})^2}\sqrt{\sum_i^N (x_i^P - \bar{x_i^P})^2}} \tag{5}$$

where $\mathbf{x}_S = (x^s)$ is the true stimulus position and $\mathbf{x}_P = (x^P)$ is the predicted position.

2.7 Simulation Study

To investigate how the sizes of receptive fields in the target regions of interest affected the decoding accuracy of x-y position, we conducted a simulation study using a voxel receptive field model. In the model, each voxel of brain activity is parameterized as having a receptive field center $\mathbf{x}_i = (x_i, y_i)$ at a "preferred stimulus position" and receptive field width (σ) indicating the voxel's tolerance to stimuli different from the preferred stimulus. Voxel responses decrease in a gaussian-like fashion as a function of the stimulus $\mathbf{x}_s = (x_s, y_s)$ distance from the RF center:

$$r_i = \exp\left(\frac{|\mathbf{x}_i - \mathbf{x}_S|^2}{2\sigma_i^2}\right) + \varepsilon \tag{6}$$

$$p(\varepsilon|\sigma_N) = N(0, \sigma_N) \tag{7}$$

where r_i is the ith voxel's response, $\mathbf{x}_i$ is the position of the voxel's receptive field center, $\mathbf{x}_S$ is the position of the stimulus, σ_i is the receptive field width, and ε is a gaussian noise term, independent and identically distributed for each voxel.

This model is a direct extension of similar models for neurons with the exception that instead of assuming a poisson process for the output, gaussian noise is used, which may be more appropriate for fMRI activities, where the sum of a large number of random noise variables (scanner noise, variation in subject cognition, and physiological changes, among others) will converge towards a Gaussian distribution by the central limit theorem. The assumptions that voxels, which are thought to reflect activity from a population of many neurons on a mm-scale, can be modeled similarly to neurons, has been supported by several recent works [17, 18]. To implement the model, a series of randomly placed targets were generated at positions (x_T, y_T) and the responses calculated for $25 \times 25 = 625$ voxels with uniformly distributed RF centers within a space of 10×10 spatial units. While changing the value of σ from 1 to 10 in 0.2 unit steps, the amount of position information in the model was quantified in two ways. First, the vector of responses from all voxels was used as input to train a sprase linear regression model with target vectors given by the true target positions. A 10-fold cross-validation procedure was used and the error evaluated using the normalized error given in equation 5. Second, a maximum likelihood solution for the position of the stimulus given the response vector and model parameters was calculated by maximizing the log-likelihood of the response according to equation 9:

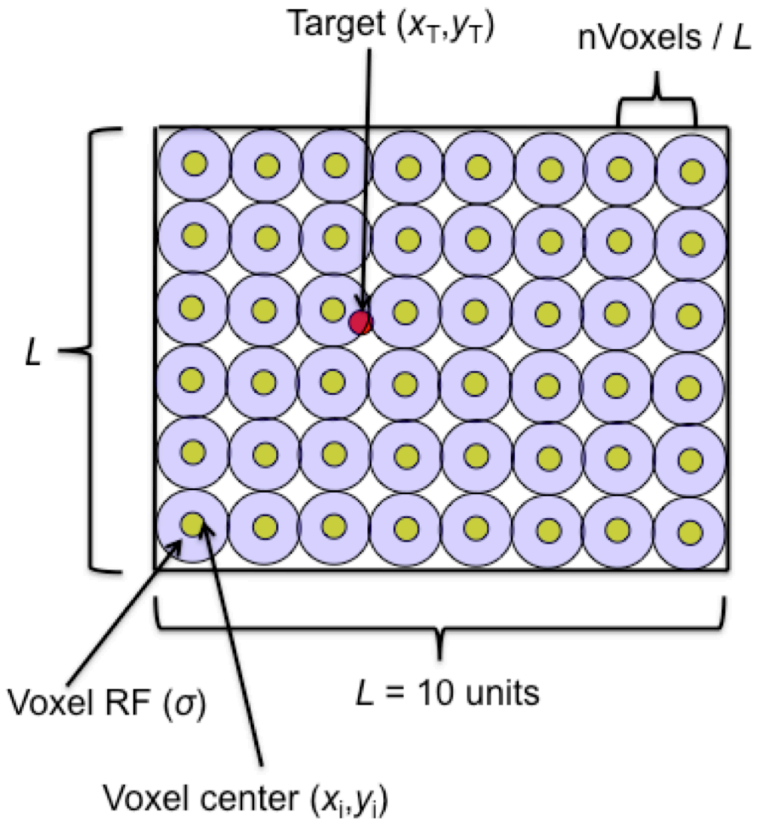

Figure 4. Voxel Receptive Field Model

A 2-dimensional area of visual space is represented by the bounds of the black box. Each small yellow circle represents the center of a voxel receptive field, and the surrounding purple circle illustrates the partial extent of the field. The red circle represents a target stimulus, to which the surrounding voxels will respond according to equation 6.

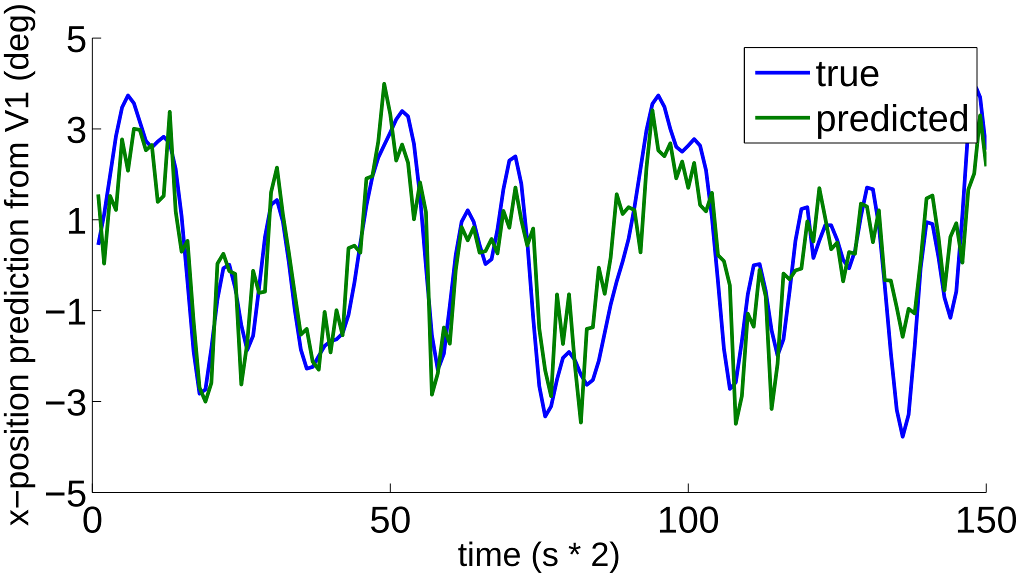

Figure 5. Sample predicted time course for x-position decoding

True x-trajectory shown in blue, predicted trajectory shown in green. A timecourse for 150 samples is plotted on the x-axis, and the position of the sphere as distance in degrees from the center of the screen is shown on the y-axis.

$$p(\mathbf{r}|\mathbf{x}_s) = \prod_{i=1}^{N} p(r_i|\mathbf{x}_s) = \prod_{i=1}^{N} \exp\left(-\frac{(r_i - f_i(\mathbf{x}_s)^2}{2\sigma_N}\right) \tag{8}$$

$$\log p(\mathbf{r}|\mathbf{x}_s) = \sum_{i=1}^{N} \left(-\frac{(r_i - f_i(\mathbf{x}_s)^2}{2\sigma_N}\right) \tag{9}$$

where $f_i(\mathbf{x})$ represents each voxel's tuning curve as given in equation 6.

3. Results

3.1 Results of Decoding Analysis

The results of the 4-fold cross-validation decoding analysis for two subjects are presented in Fig. 6. The correlation coefficient (equation 5) between true and predicted trajectories for x and y models are averaged together, and organized by visual area used (V1, V2, V3, IPS, and LOC). In addition, the results for a "dummy decoder" trained using fMRI samples whose labels were randomly shuffled is included for comparison (shown as "con" in Figure 6). The results show that within early visual cortex (V1 to V3), increasing stage in the visual hierarchy is associated with increased performance, with the average correlation coefficient r over both dimensions values increasing from 0.80 to 0.88 across the three areas (average of two subjects). A sample predicted time course in one dimension (x position) is shown to illustrate how the prediction results are correctly capturing the trend of the object movement. (Fig. 5). Reasonably high performance was also observed in higher visual areas (both LOC & IPS), with the average r values equaling 0.72 for LOC and 0.68 for IPS. In comparison, the control condition yielded an average r of 0.08.

3.2 Results of Simulation Study

A voxel receptive field model was constructed in order to theoretically assess the effect of receptive field size on the accuracy of spatial information encoding. A 10 × 10 space was uniformly covered with voxel receptive fields with gaussian tuning (see see section 2.7) and additive gaussian noise, and the responses to randomly placed stimuli used to estimate the true position of the stimulus. The accuracy of the model was evaluated in two ways. For comparison with experimental results, a sparse linear regression model was trained using the model voxel's responses to 100 randomly generated target stimuli (see section 2.7) and evaluated on a separate set of 10 samples. Secondly, the maximum likelihood estimation of the stimulus position from the model response was calculated in order to obtain an estimate of total information encoded that was independent of the choice of decoder used. Fig. 7 shows a graph of the correlation coefficient between the true and predicted values of the stimulus as a function of the receptive field widths used for each voxel. It can be seen that in agreement with the our experimental results, for the linear regression case higher accuracy was obtained as the size of the receptive field increased. However, the maximum likelihood method yielded results that were entirely independent of the receptive field width σ, implying that the actual amount of spatial information encoded in each visual area should not change, under the assumption that receptive field distributions are the same. The reason and implications behind this result are discussed in section 4.1.

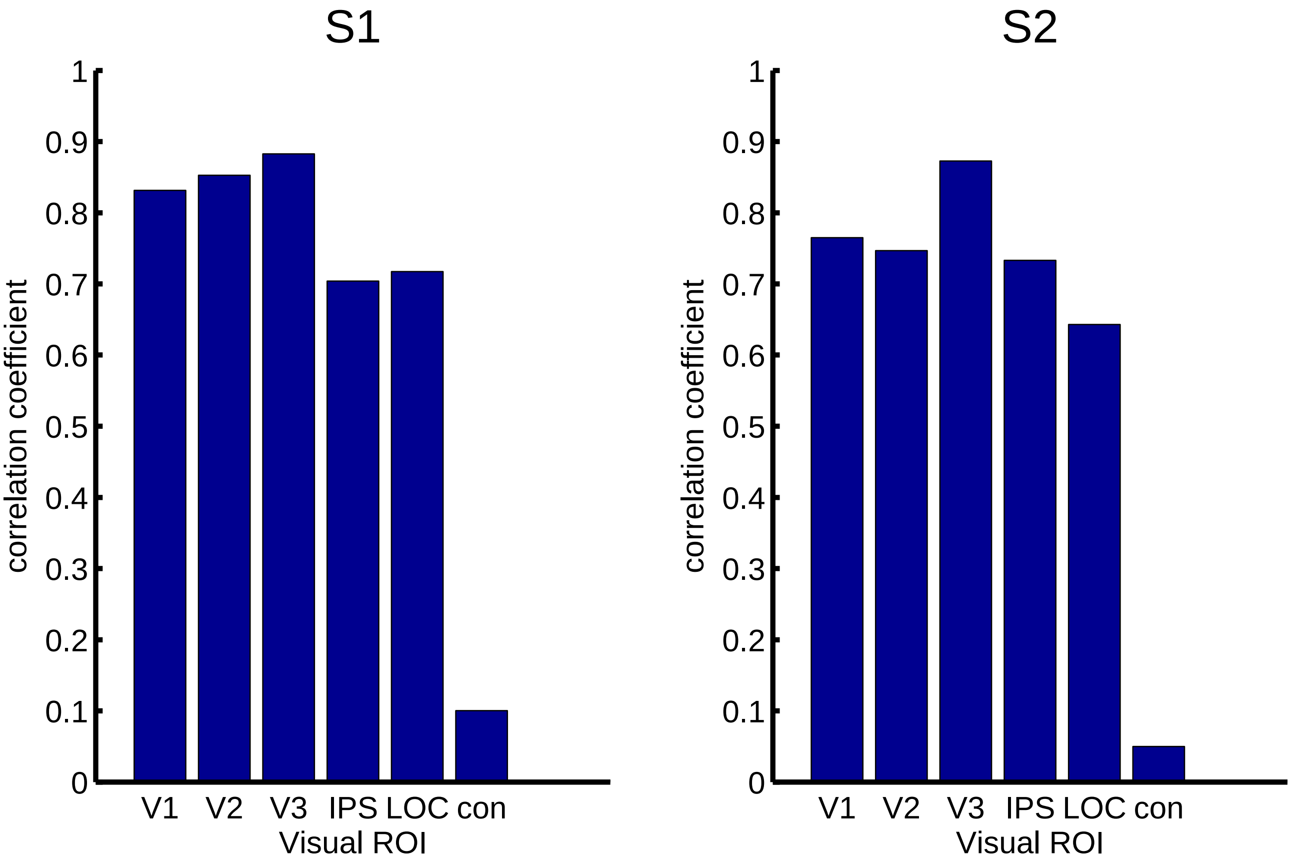

Figure 6. Error between true and predicted trajectories by visual area.

The correlation coefficient (r) between true and predicted x and y trajectories is averaged and shown on the y axis. The visual areas whose voxels were used as input for the models are arranged on the x-axis. con: control condition in which sample labels for training were randomly shuffled. S1: subject 1, S2: subject 2.

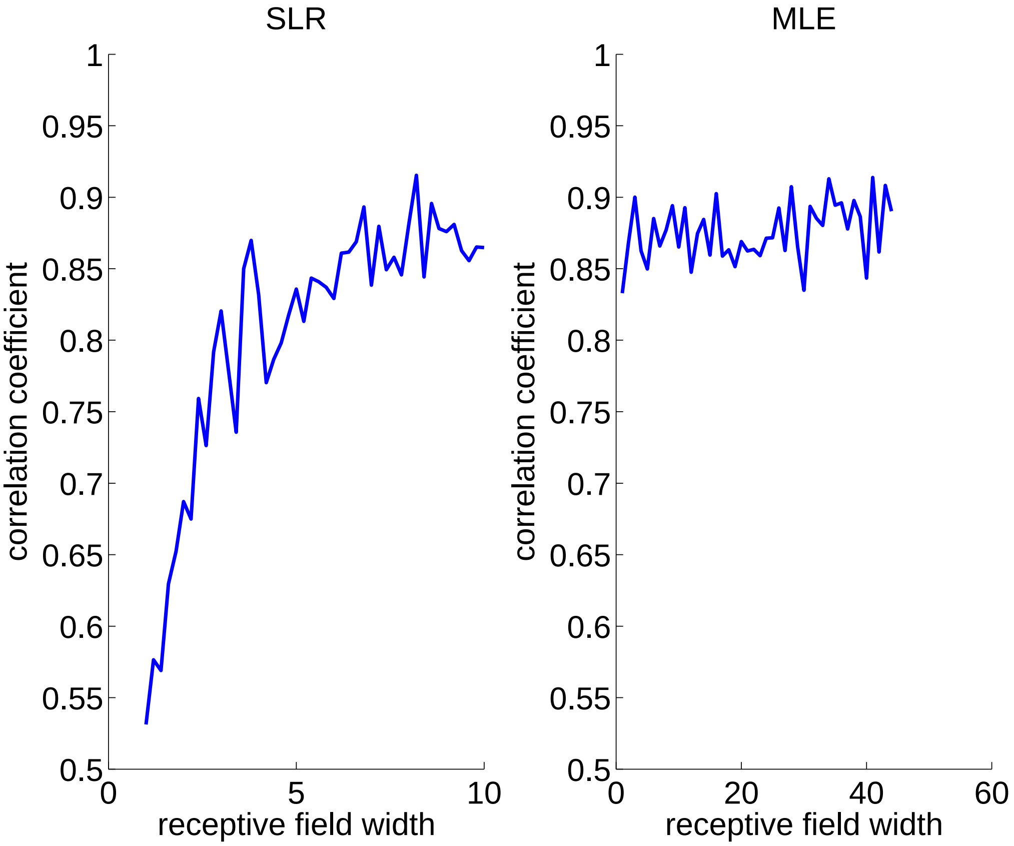

Figure 7. estimation accuracy as a function of RF width.

The correlation coefficient between true and predicted stimulus locations for the simulation study is shown on the y-axis. The value of sigma is plotted on the x-axis. Left: simulation using a sparse linear regression model. Right: simulation using a maximum likelihood estimation method (see main text for more details).

4. Discussion

The accuracy of the encoding of spatial position information in human visual cortex was experimentally explored by decoding the position of an observed moving object from human fMRI data. By training decoders with brain activity from different visual areas, we were able to compare the accuracy of spatial encoding across the visual hierarchy, and can examine the relationship between accuracy and known properties of the visual areas.
For example, it is well known that the responses of neurons in higher visual areas are more complex than of those in lower visual areas, and likewise have larger receptive fields to incorporate information from a wider range. In this discussion we will consider our decoding results with respect to these aspects separately.

4.1 Size of Receptive Field by Visual Area

Previous studies have shown that the sizes of receptive fields tend to increase with successive stages in the visual processing hierarchy[11]. Because intuitively it may seem that using smaller receptive fields would allow for a more precise encoding of spatial information, the result (6) that accurate position prediction was possible from all areas was initially surprising. However, the results of the simulation study (Fig. 7) and previous theoretical research [12] has shown that the accuracy of encoding or decoding in 2 dimensions is not dependent on the size of the receptive fields. In [12] it was shown theoretically that there is a universal scaling rule for receptive field width, meaning that regardless of the probabilistic model or exact form of tuning function assumed for neuronal receptive fields, there is a specific way in which the change in size affects encoding accuracy. Importantly, this rule depends on the dimensionality of the variables being encoded: for 1-dimensional variables, increasing the RF size will always lower the encoding accuracy, for 2-dimensional variables the size has no effect, and for 3-and greater dimensional variables, larger RF sizes lead to better encoding accuracy (8).
Thus, despite the much larger receptive field sizes in areas such as LOC or IPS [11], adequate accuracy was obtainable in this 2-dimensional encoding setting. The pattern of increasing accuracy from V1 to V3 could be a result of the choice of regression method, which was linear as opposed to the maximum likelihood

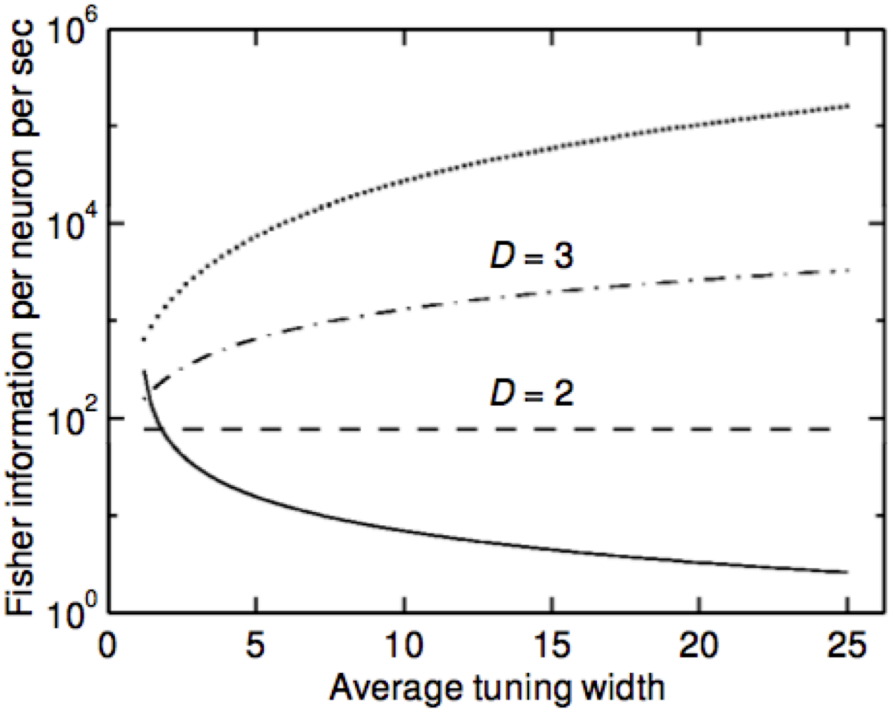

Figure 8. Fisher information for neuronal populations as a function of tuning curve width.

Adapted from [12]. Depending on the dimensionality D of the encoded variable, the Fisher information for the neuronal population can decrease, increase, or be unaffected by an increase in average tuning width.

method used for the simulation study. A future task is to apply the maximum likelihood framework to actual fMRI data and observe whether the results by area truly do not differ.

4.2 Level of Retinotopy by Visual Area

Areas V1 through V3 have very well-defined retinotopies, meaning that every point of visual space is adequately covered by neuronal receptive fields and that adjacent RFs in the visual field belong to neurons adjacent in the cortical sheet. However higher visual areas such as LOC, which is dedicated to creating invariant representations of stimuli, do not have similarly 'clean' retinotopies [21]. This distinction is important when considering the theoretical implications of receptive field size on encoding accuracy because typically the assumption of densely packed, uniformly distributed receptive fields is necessary for deriving the independence of RF size from encoding accuracy [12]. This could explain why areas LOC and IPS showed a higher error for decoded trajectories despite the rationale that their large receptive field sizes should not hinder spatial encoding.

5. Conclusions and Future Work

The accuracy of the position encoding of objects across different visual cortical areas was explored in this thesis. The increase in receptive field size across areas was also related to the increase in accuracy for early visual cortical areas, although the complexity of responses in higher visual cortex resulted in lower accuracy compared to early visual cortex. The dorsal visual stream, which is posited with representing object position for enabling human interaction is a complex visual area which may encode position in a manner fundamentally different from the retinotopic way in which early visual cortex encodes this information. Thus, a more complete survey of the representation of object position in dorsal visual cortex is a problem left for future works. In addition, the results from the simulation study imply that the level of position information does not change with receptive field size, but this trend was not confirmed when using a linear decoder on the fMRI Data. An important point for consolidating this work is to estimate the sizes and positions of the receptive fields manually (following [1]) and then attempt a maximum likelihood estimation of the positions and observe whether the theoretical results are supported.

Acknowledgements

There are many people I would like to thank for their help and support with my research. First and foremost, I would like to thank my direct supervisor Yukiyasu Kamitani who collaborated closely with me and gave me overarching guidance and helped with many of the conceptual works for this thesis. For help with experiments, critical analysis of my work and support in all things work-related, I would like to thank Tomoyasu Horikawa. For giving inspiration, leadership, and much good advice I would like to thank Yoichi Miyawaki. For encouragement and interesting conversations I thank Masako Tamaki. For being so incredibly reliable and always helping no matter the circumstance, I thank Professor Kazushi Ikeda. For lending many useful comments and help with some mathematical aspects of my research I thank Kei Majima. In addition, I would like to thank all others who are part of DNI, ATR, and NAIST who have supported me and given me the ability to make it this far. Thank you all!

References

[1] Wandell B.A., Dumoulin S.O., Brewer A.A. (2007). Visual Field Maps in Human Cortex. Neuron 56,

[2] Carandini M., Demb J.B., Mante V., Tolhurst D.J., Dan Y., Olshausen B.A., Gallant J.L., Rust N.C., 2005. Do we know what the early visual system does? J. Neurosci. 16, 10577-97.

[3] Hubel, D. H. & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. Journal of Physiology, 195, 215-243.

[4] Kanwisher N, McDermott J, Chun M. M (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17: 4302?4311.

[5] Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. Nature 392: 598?601.

[6] Aguirre G. K, Zarahn E, D'Esposito M (1998) An area within human ventral cortex sensitive to 'building' stimuli: evidence and implications. Neuron 21: 373?383.

[7] Ronald S. Fishman (1997). Gordon Holmes, the cortical retina, and the wounds of war. The seventh Charles B. Snyder Lecture Documenta Ophthalmologica 93: 9-28, 1997.

[8] Ungerleider, L.G. & Haxby, J.V. 'What' and 'where' in the human brain. Curr. Opin. Neurobiol. 4, 157?165 (1994).

[9] Goodale M.A., and Milner A.D. (1992). Separate visual pathways for perception and action. Trends in Neurosciences, 15(1):20?25, 1992.

[10] Schwarzlose R.F., Swisher J.D., Dang S., Kanwisher N., 2008. The distribution of category and location information across object-selective regions in human visual cortex, Proc. Nat. Acad. Sci. 105, 4447-4452

[11] Dumoulin, S.O. & Wandell, B.A. Population receptive field estimates in human visual cortex. Neuroimage 39, 647?660 (2008).

[12] Zhang K., Sejownki T.J., 1999. Neuronal Tuning: To Sharpen or Broaden?, Neural Computation 11, 75-84

[13] Yamashita O., Sato M., Yoshioka T., Tong F., Kamitani Y., 2008. Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns, NeuroImage 42, 1414-1429

[14] Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., and Tootell, R.B.H. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268, 889?893.

[15] Culham, J.C., Brandt, S.A., Cavanagh, P., Kanwisher, N.G., Dale, A.M., & Tootell, R.B.H. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. Journal of Neurophysiology, 80, 2657-2670.

[16] Bishop C.M. (2006). Pattern Recognition and Machine Learning. Springer.

[17] Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8, 679?685.

[18] Nishimoto S., Vu A.T., Naselaris T., Benjamini Y., Yu B., Gallant J.L., 2011. Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. Curr. Biol., doi:10.1016/j.cub.2011.08.031

[19] Swisher D.J., Halko M.A., Merabet L.B., McMains S.A., Somers D.C., 2007. Visual topography of human intraparietal sulcus. J. Neurosci. 27, 5326-5337

[20] Inouye, T. (1909). Die Sehstroungen bei Schussverietzungen der kortikalen Sehsphare (Leipzig, W.: Engelmann).

[21] Rosa M.G.P. (2002). Visual maps in the adult primate cerebral cortex: some implications for brain development and evolution. Braz J Med Biol Res, December 2002, Volume 35(12) 1485-1498

[22] http://www.cse.buffalo.edu/ rapaport/575/vision

Appendix

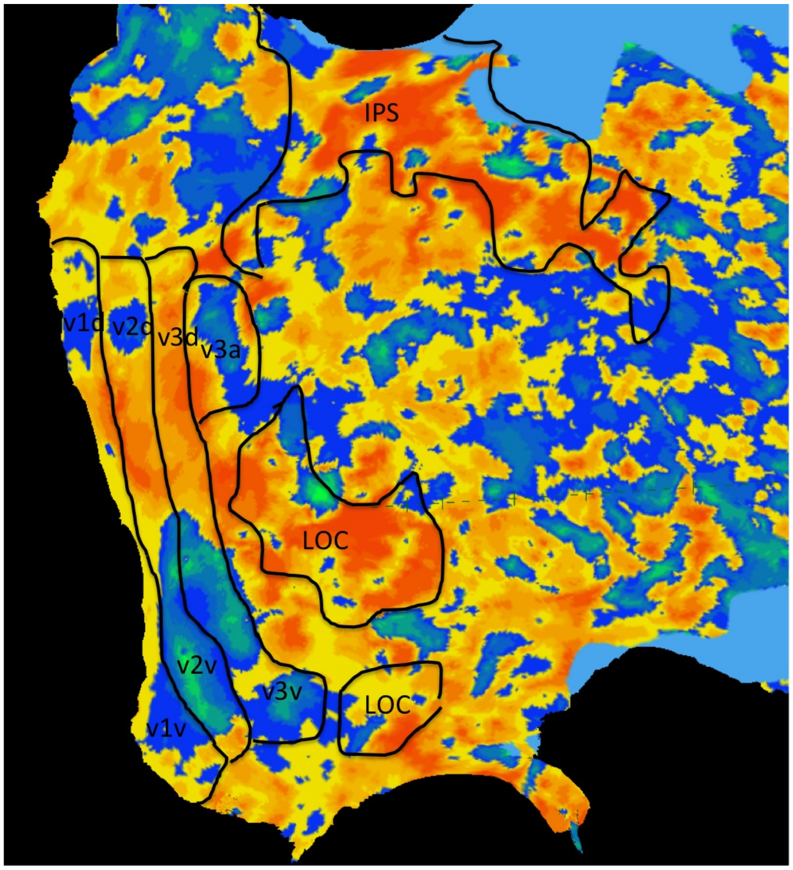

Figure 9. Definitions of visual areas.The retinotopic map for the right hemisphere of subject 1 is displayed here. It was created using BrainVoyager QX and the retinotopic stimuli described in [14]. The borders of visual areas were defined by the phase-reversal boundary for areas V1, V2, and V3, and by functional localizer experiments ([15],[4]) for areas LOC and IPS.