RESOLVING THE PROJECTION OF AN OCCLUDED STIMULUS ON THE HUMAN CORTICAL SURFACE

Kevin DeSimone

A THESIS SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ARTS

GRADUATE PROGRAM IN PSYCHOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO

FEBRUARY 2013

Abstract

The human visual system is capable of tracking multiple visual targets under a variety of task constraints and configurations. For nearly two decades, the psychophysical literature has shown that moving, occluded visual targets—targets that are momentarily invisible as they pass behind an occluding bar—are differentially represented by the visual system compared to their moving, non-occluded counterparts. Here, I sought to examine the neurophysiological basis of this behavioral difference in response to occluded versus non-occluded visual targets. I used brain imaging to conduct modern retinotopic mapping experiments in human participants. Once their early visual cortices were mapped, I was able characterize the neural representations for both targets and distractors as well as during moments of occlusion and non-occlusion. The results show that, using our method, we can distinguish visual targets from distractors; furthermore, there appears to be a representation in retinotopically organized early visual cortex for visual targets that have momentarily disappeared from the visual field due to occlusion.

Table of Contents

List of Figures

1. Introduction

This thesis project started out as a small enterprise aimed at determining how the brain represents a moving visual stimulus during a tracking task. The scope of the project was initially circumscribed so as to explain how early visual cortex differentiates between targets and distractors within the context of its representational architecture. Over the course of two years, the project has grown into a package of somewhat diverse analytic approaches and experimental findings. To answer questions about the dynamic representation of a moving object across multiple visual areas, I drew upon methods and findings from psychophysical, electrophysiology, and neuroimaging. The thesis is organized to gradually build up a case for using novel neuroimaging techniques to determine the role that multiple areas in early visual cortex play in extracting and representing visual information during a tracking task.

Visual tracking is an interesting psychological and behavioral state from the perspective of neuroimaging, especially in early visual cortex, because it represents the intersection of bottom-up and top-down visual processing. Different visual stimuli of equivalent physical properties can be assigned the role of target and distractor arbitrarily by the experimenter. From trial to trial, participants have no trouble maintaining this distinction and can perform tracking tasks made especially difficult with very little decrement in their performance. This scenario begs the question: how do top-down, attention-driven task demands affect the bottom-up, stimulus-driven representations of visual stimuli in

early visual cortex? The answer to this question has been put forth numerous times in the literature in the past decade, but none have proposed methods or reported findings that differentiate sensory-bound signals from attention-driven signals in early visual cortex. This absence is in part due to what seems like an analytic and experimental design oversight of one of the most fundamental properties of early visual cortex: retinotopic organization.

After describing some basic methodological details that subserve all experimental aspects of the thesis, I summarize the development of retinotopic mapping in visual cortex, from nascence to the state of the art. Recent advances in neuroimaging has moved towards an approach that attempts to describe the point-for-point registry between visual and cortical space with a level of precision suitable for dissociating concurrent visual processes deployed across the visual field. Such an approach offers insight into how bottom-up, stimulus-driven responses might differ from top-down, attention-drive signals across multiple visual areas and for different task conditions.

The fourth section presents a novel neural decoding method in the field of neuroimaging. Here, I detail the development of an algorithm for generating reconstructions of visual stimuli based on the retinotopic mapping and functional time-series acquired with functional brain imaging. The method allows experimenters explore aspects of how the brain's voxel-wise activity varies as a function of the pattern of visual stimulation and task instruction. Rather than claiming something about how a given visual area changes its response during

tacking compared to rest, we are able to ask questions about how a given visual area represents a portion of the visual space while a target passes through it compared to a distractor.

The fifth section of the thesis goes on to implement the stimulus reconstruction method for a scaled-down version of a tracking task. The method allows for the teasing apart of signals associated with tracked and non-tracked items across multiple visual areas. I was able to demonstrate that the signal associated with tracked and non-tracked items as well as occluded and non-occluded items can be differentiated within the same visual area at the same moment in different retinotopic locations.

The final section to the paper attempts to make sense of the results of the tracking experiment both in terms of the psychophysical tracking literature as well as the underlying neurophysiological meaning of the recorded brain imaging signal. The story for what the expected brain activity should be given the imaging technique and the task demands is quite complex. I try to frame this complexity in terms of how the brain activity in early visual cortex might express forms of stimulus enhancement and/or inhibition as a result visual stimulation and task demands. I offer several possible explanations for my results, as well as some suggestions for experiments moving forward.

1.1 Multiple-object tracking

The multiple object tracking (MOT) paradigm was first introduced by Pylyshyn and Storm (1988). Myriad variations on the original MOT task structure have en developed throughout the years, but the essential spirit of the task has remained relatively the same. Participants fixate a central point on a visual display while an array of dots are presented on the screen at the outset of a trial. A subset of the dots are indicated as targets, usually via a brief change in luminance. Following the cue period, all of the dots set off in motion at a constant velocity around the display space. After some time, the dots stop moving and a single dot is probed. Participants are required to respond as to whether the probed dot is a target or distractor.

Storm and Pylyshyn (1988) discovered that participants performed quite well at distinguishing targets from distractors for up to target and distractor set sizes of 10, performing about 85% correct; tracking more than 5 targets (or contrarily ignoring more than 5 distractors), was shown to be beyond the ability of the participants (Storm & Pylyshyn, 1988; Pylyshyn, 1989). This pattern of performance as a function of target set size has been replicated across the years for various MOT stimuli aside from dots, including surfaces (vanMarle & Scholl, 2003), objects (Alvarez & Franconeri, 2007), and faces (Ren et al., 2009). Since targets and distractors were identical in terms of their featural qualities, Storm and Pylyshyn reasoned that the visual processes that support the participants'

tracking performance must individuate attended and ignored stimuli based on their current and historical spatial locations.

Storm and Pylyshyn constructed a model to determine whether participants' performance on the MOT task for target set sizes of up to 5 items relied on parallel or serial tracking mechanisms in the visual system. They rightly pointed out that if the mechanisms of visual tracking were truly operating in parallel, then there should be no gradient of performance as the number of tracked (or ignored) items increases. Conversely, If visual tracking operated via a serial mechanism whereby participants rapidly shift the locus of attention around the visual field from item to item in rapid succession in order to maintain the identifies of targets and distractors, then errors and latency should increase with both the number of targets and distractors as well as the distance between them.

The experimenters estimated the trade off between attentional velocity (i.e., the rapidity of attentional shifts in a serial tracking process) and performance and found that participants' performance for target set sizes of up to 5 items outpaced the performance of a serial tracking model (Pylyshyn, 1989). With target set sizes larger than 5, participants' performance drop off precipitously. That participants outperform a serial tracking model but show a degradation in performance with large target set sizes indicates that tracking sits somewhere between a purely serial and purely parallel process. That is, for a certain number of items to be tracked, the visual system can deploy tracking

mechanisms for each target in parallel; however, with target set sizes beyond this limit, the processing demands exceed the capabilities of parallel deployment. Pylyshyn hypothesized the existence of "fingers of instantiation" (FINSTs) to explain how multiple targets can be individuated automatically and in parallel to subserve processes engaged in the MOT task. A FINST is resource-limited mechanism that individuates features in the visual scene, but which is separate from its retinal location per se. In the case of the classic MOT experiment where targets and distractors are identical in luminance and form, their feature-based distinguishability relies solely on their spatial locations and histories. Hence, spatial location becomes the feature indexed by FINSTs associated with each target, allowing the visual system to correctly identify a target in motion over time.

The individuation of targets via the FINST mechanism is thought to occur in a pre-attentive stage of processing, analogous to the feature integration model proposed by Treisman and Galade (1980) to explain various behavioral phenomena of participants engaged in a visual search paradigm. Visual search comes in two basic flavors. Feature search where targets and distractors differ in terms of a single feature demonstrates the pop-out effect. Under such a search regime, the reaction time for finding the target does not scale with the number of items in the array. Instead, the target will effortlessly pop out because the feature has been abstracted in the preattentive stage. Conjunction search, on the other hand, uses search arrays where targets and distractors share two or more features. Reaction times will scale with the number of items in the array as

attention is guided by the conjunction of two or more preattentive features, behaviorally manifested as participants conducting a serial search through the array. Efficiency of search relates the speed at which participants are able to identify the target and the number of items in an array. Search efficiency is used to delineate separable features from integral features. That is, if a particular search array is found to be inefficient, then it is reasoned that attention is required to bind multiple separable features into an integral feature as the participant searches for the target among distractors. Efficient searches indicate that preattentive processes select on a separable feature and guide attention to the target directly. The two-stage model was put forth as a way of explaining how certain conjunctions of features seem to require a serial search strategy and longer reaction times whereas searches for the constituents of those conjunctions do not. The model sets out a framework for experimentally determining, for a given scene, what should be considered a separable or integral feature.

Within the FINST model, spatial location becomes a feature through which multiple items can be individuated, tracked, and identified when probed. An interesting extension of the classic MOT paradigm has shown that participants' performance can be modified when the targets and distractors interact with occluding bars in the display. Scholl and Pylyshyn (1999) found that participants' were able to track targets regardless of whether the targets passed behind occluders. In other words, even though the stimulus is momentarily extinguished

from the display, participants have no difficulties in persisting the representation (FINST) long enough for the stimulus to traverse the occluder and reappear on the display. Interestingly, the experimenters also found that this equivalence between occluded and non-occluded tracking can be destroyed when unnatural forms of occlusion are used. In the natural world, objects that pass behind and in front of one another cause their boundaries to delete and accrete in an ordered succession. Scholl and Pylyshyn designed control conditions where occlusion occurred in an unnatural way. Here, the objects would implode and explode or instantaneously disappear and reappear upon intersecting and reemerging from behind an occluder. In these cases, participants were up to 30% worse at the MOT task with unnatural occlusion compared to natural occlusion. These findings imply that the visual system not only represents occluded items but that there may be some mechanisms that cannot identify unnatural occlusion in the course of deploying tracking resources.

Extending these findings, Flombaum and others (2008) found that the deployment of attentional resources can be biased by occlusion. Here, they had participants track targets and ignore distractors transiting around a field with two occluding bars. Participants were able to perform the tracking task equally during occlusion and non-occlusion trials, successfully identifying targets around 90% of trials. To investigate how attention varies as function of occlusion and tracked item identity. the experimenters included the appearance of brief (100 ms) probes that the participants were asked to acknowledge via button response.

Targets and distactors were probed during moments of occlusion and nonocclusion an equal number of times through all trials. Two interesting findings fell out of the behavioral responses to probes among these four conditions. First, they confirmed the existence of distractor inhibition described previously (Pylyshyn, 2006). Distractor inhibition manifests itself behaviorally as a lower probe detection rate (20%) for probes that occur at or near distractors compared to targets. This finding suggests that attention deploys some kind of mobile suppressive resource that has information about the location and identity of a particular tracked item.

In addition to distractor inhibition, Flombaum and others posited a new behavioral finding they dubbed the "attentional high-beams effect". The attentional high-beams effect manifests itself as a behavioral advantage for identifying target *and distractor* probes during moments of occlusion compared to nonocclusion. Specifically, participants were 22% and 33% better at identifying probes on targets and distractors during occlusion compared to nonocclusion. In other words, participants are better at the detecting probes near targets and distractors during moments where occluders obscure them from view. Interestingly, distractor probe detection benefits more from occlusion than does target probe detection; however, target probe detection was always higher than distractor probe detection.

Distractor inhibition and the attentional high-beams effect offer two interesting behavioral cases for investigating the role of neural mechanisms that

support visual tracking specifically and perhaps visual spatial attention generally. How do the neural representations of targets and distractors differ in light of distractor inhibition? How do the neural representations of occluded and nonoccluded tracked items differ in light of the attentional high-beams effect? Surprisingly little is known about the answers to these questions, despite the leaps and bounds made in the past decade in the field of human brain imaging. To position ourselves better to answer the questions about the underlying neural mechanisms supporting tracking and its various quirks, a review of the brain imaging literature focusing on multiple object tracking is helpful.

1.2 Neural correlates of tracking

In comparison to the vast wealth of psychophysical experiments employing variants of the MOT task, relatively little is known about the neural correlates of MOT, and the neural representation of complex stimuli in general. The majority of studies using functional brain imaging (fMRI) to measure brain activity related to tracking tasks rely on statistical methods to detect regions of the brain that show elevated blood-oxygenation-dependent (BOLD) signal for tracking versus passive-viewing blocks of trials. The experimental paradigm is an alternating task structure whereby participants' brain activity are recorded during epochs of effortful tracking and during passive viewing. Statistical maps are computed that summarize the pattern of activity among all the voxels during each of these task

epochs. The activity that occurs during tracking above and beyond the baseline activity is computed as the difference between the two task states.

In the first experiment to probe the underlying neural mechanisms supporting visual tracking, Culham and others (1998)—using the subtraction method—found that compared to a passive viewing condition, voxels throughout occipitotemporal, parietal, and frontal cortices showed significant activation across both hemispheres during a tracking task. In particular, they found that both the motion-sensitive MT+ and the form processing lateral occipital cortex (LOC) showed significant activation for tracking versus passive viewing. in addition, they found multiple foci of significant activation in parietal cortex for the tracking versus passive viewing contrast, including two sites in the intraparietal sulcus (IPS) and superior parietal lobule (SPL). In the frontal cortex, they found that frontal eye fields (FEF) and supplementary motor area (SMA) were significantly more active for tracking versus passive viewing (Culham et al., 1998).

A major problem with this study, and with the statistical contrast method in general, is the lack of specificity in the contrast between an effortful task state and passive viewing. Tracking as a psychological and behavioral state brings with it an array of neural mechanisms that subserve the task demands. Passively viewing a MOT task may crudely control for the low-level visual components of a tracking task; however, this condition would not account for the increased demands on spatial visual attention, vigilance, or response planning

among others. The contrast method lumps together all higher-level perceptual and cognitive processes into a monolithic statistical effect in contrast to passive viewing. To more finely gauge the involvement of different task demands and their corresponding neural mechanisms, experimenters have used the method of parametric variation to define the involvement of different brain regions in the MOT task.

Rather than simply defining two states of the task and the corresponding brain imaging data, the parametric variation approach is to use small, incremental changes in the task demands in order to encode corollary incremental modulations in the BOLD activity among voxels measuring neural activity related to the task. Within the realm of MOT, this is usually implemented by way of increasing the number of items to be tracked. With increased numbers of targets, the supposition would be that BOLD activity should scale accordingly in areas that are specifically representing tracked items. Any brain areas which may be active to support a general vigilance or attention component of the task that does not scale with the number of tracked items should not change its level of activation.

In a follow-up to their original fMRI tracking experiment, Culham and others (2001) found that the pattern of results in their earlier work was not as straightforward as their simple contrasts suggested. Here, the participants tracked between 0 and 5 targets, as indicated at the outset of a given trial. In this context, 0 targets is analogous to the passive viewing condition discussed earlier

in relation to their original 1998 experiment. The tracking component of the task was only present for target set sizes of 1 and greater. The parametric design allowed the experimenters to parse out patterns of activation driven by the general demand characteristics of the task from the patterns of activation involved in tracking on the basis of whether voxels demonstrated a scaled BOLD amplitude in conjunction with variable target set size. They found that sites in IPS—specifically the anterior intraparietal sulcus (AIP) and posterior intraparietal sulcus (PIP)—showed increased BOLD amplitude that scaled with the number of items tracked. In contrast to their original 1998 findings, while voxels in areas FEF, SPL and MT+ all showed significantly greater activation for tracking versus passive viewing (i.e., 0 targets), the activation levels in these areas did not parametrically vary with the number of distractors. In other words, the BOLD amplitude in these three brain areas remained tonically high during periods of tracking regardless of the number of items participants were cued to track. This seems to suggest that while certain areas in IPS subserve specific target-related demands of the tracking task, areas FEF, SPL, and MT+ are recruited to support non-specific task demands such as spatial attention, vigilance, and response planning (Culham et al., 2001). These results have been replicated in other functional imaging studies using the parametric variation of attentional load during a tracking task (Jovicich, 2001; Howe, 2009). In general, the level of activation of voxels in posterior parietal cortex seems to scale with the attentional load of the tracking task; frontal areas, including FEF and SMA, do not show

patterns of activation that scale with target set size. These findings are suggested to demonstrate the role of parietal cortex in the deployment of visual spatial attention resources.

These findings add detail to the body of literature showing broad task-driven responses in parietal and frontal cortices (Kanwisher & Wojciulik, 2000; Wojciulik & Kanwisher, 1999). However, a glaring absence in these imaging experiments exploring the neural representations of MOT is the pattern of activation in early visual areas. Since early visual areas are retinotopically organized and the functional topographies of these areas are well defined (Wandell et a., 2007), differences in the representation between targets and distractors should be detectable and ascribed in the coordinates of the retinotopic space. Rather than comparing the relative differences among groupings of voxels and variations of task condition, it should be possible to trace the history of a target or distractor in terms of BOLD signal modulation across the human cortical surface in retinotopic space. If such an approach is possible, then it becomes relatively simple to compare the representations of targets and distractors or stimulus and attention dynamically as a task unfolds.

1.3 Hypotheses: squaring visual tracking and functional imaging

The results of psychophysical experiments investigating the deployment of attentional resources during a tracking task suggest two interesting behavioral mainstays: 1) distractor inhibition and 2) the attentional-high beams effect.

Distractor inhibition describes the suppression of attentional salience at or near the visuotopic locations of distracting items. The attentional-high beams effect describes the enhancement of attentional salience at or near the visuotopic locations of occluded items, target or distractor. From these two behavioral findings, it is possible to begin to formulate hypotheses for how neural mechanisms may deploy resources across retinotopic space under various circumstances. Distractor inhibition should manifest itself neurophysiologically as a relative increase in blood-oxygen-level-dependent (BOLD) signal in brain tissue representing the retinotopic coordinates at or near distractors. Psychophysically, distractor inhibition can be thought of as a negative bias in the salience map of the visual field at or near distractors, yielding low probe detection rates. However, since the BOLD signal cannot dissociate excitation from inhibition (Logothetis, 2008), the expected result would be an increase in the BOLD signal for neural tissue corresponding to the location of the distractor in retinotopic space, assuming an equivalence between psychophysical and neural inhibition. The attentional high-beams effect should manifest itself neurophysiologically as a relative increase in the BOLD signal in tissue representing the retinotopic location at or near occluded items.

2. *General Methods*

The sections that follow outline the collection and analysis procedures that are common among the visual mapping and attention experiments. Special attention will be paid to describing the analyses specific to each of the experiments in subsequent chapters.

2.1 Participants

Measurements were obtained from 4 participants (two female; ages 25-31 years). All participants had normal or corrected-to-normal visual acuity and gave written informed consent approved by the York University Institutional Review Board. Participants submitted to a visual field mapping and a multiple object tracking experiment.

2.2 Magnetic resonance imaging

Magnetic resonance images were acquired with a 3T Siemens Trio Tim scanner and a 32-channel head coil. Foam padding was used to minimize the participants' head-motion. Participants were not naive to a scanning environment, having been previously and repeatedly instructed on how to comport themselves in the scanner so as to minimize head movements.

Functional images (flip angle = 90°, TE = 28 ms, TR = 1500 ms, 256 mm FOV, 128x128 matrix, 2x2x2 mm voxel, 6/8 partial phase Fourier, GRAPPA parallel imaging with 2x acceleration factor, bandwidth = 752 Hz/px) were

acquired using an echo-planar pulse sequence with 20 coronal slices oriented perpendicular to the anterior-posterior axis of the calcarine sulcus. Ten visual mapping runs and 10 tracking runs were collected from each subject in two scanning sessions on separate days.

In addition, a T_1 weighted three-dimensional MPRAGE sequence scan of entire head with an isotropic 1 mm^3 voxel (TR = 1900 ms, TE = 2.52 ms, 1 mm slice thickness, 256 x 256 matrix). The T_1 weighted sequence yielded a high contrast, high spatial resolution 3D structural image from which to derive surface reconstructions for functional data projections.

2.3 Image processing

All anatomical images were processed through the FreeSurfer software package (Dale et al., 1999; Fischl et al., 1999), yielding surface reconstructions onto which statistical maps can be projected. For clarity in presenting data, the surfaces can be cut and flattened, offering a single-perspective view of an entire hemisphere. Smaller sections were cut from these hemispheric flat maps, isolating the occipital lobe of each hemisphere.

All functional data were processed through a common analysis pipeline. All runs were slice-time corrected to ensure that the time-series of all voxels within a volume acquisition were aligned with respect to the visual stimulation. Data were then motion-corrected to the first volume of the first run for a given scanning session. Each run was deskulled and each time-series mean-centered

and variance normalized. Since each participant was scanned in multiple sessions in order to acquire all the necessary pRF estimation and visual tracking data, it was necessary to transform all the data to a singular, common space. We used an affine registration procedure to register all functional runs with a single high-resolution T_1 from which the surface reconstructions were generated. The registration algorithm uses a local Pearson coefficient cost function (Saad et al., 2009). This alignment prescription is designed to align functional and anatomical images based on local similarities in the images. Since the gross shape of the functional and anatomical images are quite distinct due to differences in the susceptibility artifacts and field distortion between the two pulse sequences, cost functions that use global shape to drive the alignment between these types of images often fails.

2.4 Volume censoring noise reduction

In addition to these standard forms of fMRI data preprocessing in preparation for statistical analysis, we employed noise reduction procedures for removing spurious signals from our data resulting from head-motion (Powers et al., 2012) and physiologic noise (Thomas et al., 2002).

Motion-correction is a procedure whereby series of three dimensional volumes—which are typically concatenated over time to produce a single four dimensional volume—are brought into spatial alignment with some reference volume of the same spatial resolution via an affine spatial transformation. While

there are many different implementations of motion-correction for images, the common thread throughout is an error minimization between the reference image and the motion-corrected image. The result of such a procedure is a six parameter spatial transformation that describes the translation (three degrees of freedom, units are in millimeters) and the rotation (three degrees of freedom, units are in radians) required to bring the two images into alignment. In addition, these parameters tell us *how much* our participants moved during a scanning session by combining the various parameters into a singular measure. We can, in turn, use this measure to explicitly discount certain data points due to abnormally high amounts of head-motion. The measure adopted in this particular study was framewise displacement (Power et al., 2012).

Framewise displacement expresses the instantaneous head-motion in a scalar quantity with the formula

$$FD_i = |\Delta d_{ix}| + |\Delta d_{iy}| + |\Delta d_{iz}| + |\Delta \alpha_i| + |\Delta \beta_i| + |\Delta \gamma_i|$$

where

$$\Delta d_{ix} = d_{(i-1)x} - d_{ix}$$

for a particular motion parameter. The same evaluation was used for each of the other motion parameters. The rotational displacements were converted from radians to millimeters to standardize the units across parameters. Once the framewise displacement had been computed for every volume in a given

scanning session, particular volumes were flagged as containing too much head-motion if the framewise displacement measure exceeded a predetermined exclusion criterion. The exact exclusion criterion varied from scanning session to scanning session, but typically was set so as to remove the volumes whose framewise displacement exceeded the value that demarcated the 95th percentile of all volumes. Discounted volumes were removed from the computation of the mean run for a given experiment.

2.5 Component-based noise reduction

In addition to censoring volumes based on their framewise displacement as derived from the head motion parameters, an independent component analysis (ICA)was used to identify spurious signals embedded within the time-course of each voxel. Each functional dataset is processed through MELODIC (Beckmann & Smith, 2002), a free software tool that decomposes a four dimensional brain imaging dataset into a set of independent spatial components, each of which is associated with a particular temporal component. The linear decomposition of the voxel time-series via ICA assumes that a given voxel time-course is the weighted sum of an indeterminate number of components. However, because ICA is model-free, no automated tools exist for identifying certain components as being stimulus-driven from those which owe to physiologic noise such as respiration or scanning artifacts such as susceptibility artifact. Noise components were identified if their frequency was less than half

or more than twice our stimulus frequency. For instance, the stimulus frequency of the bar stimulus was computed to be 0.067 Hz, where visually responsive voxels should show 8 peaks corresponding to the 8 bar sweeps over the course of the 120 volume run. Any ICA components which showed a frequency less than 0.033 Hz or more than 0.133 Hz were flagged as being physiologic noise or otherwise as not emanating from the visual stimulation.

Once flagged, noise components were removed from the raw time-series of each voxel using the standard multiple regression expression

$$Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} \ \ldots + \beta_1 X_{ip} + \varepsilon_i .$$

However, in our approach to reducing the contribution of noise to a given voxel time-series, we were less interested in the parameter estimate β than in the residual error ε. In the applying general linear model, each of our independent variables is a temporal component derived from the ICA which is outside of the frequency band we've defined for a stimulus-driven signal. Hence, the variability in our data that is not accounted for by the designated noise components is assumed to reflect stimulus-driven activity.

3. Mapping the visual space using functional brain imaging

To be able to ask questions about differing neural representations for targets and distractors under occlusion and nonocclusion, it is necessary to take a cartographical approach to early visual cortex. The changes in the neural representation during a tracking task will vary both in terms of the configuration of the task structure as well as the locations of the targets and distractors. In order to fully dissociate the signals related to targets and distractors or occluded and nonoccluded items, we need to precisely relate the visual space to cortical space. Defining the stimulus in visual space, then, should lead us to the corresponding signals in the cortical space. This retinotopic mapping of human cortex is not a new enterprise, and emanated from somewhat humble beginnings.

3.1 Retinotopic Mapping in the Human Visual System

The retinotopic organization of human occipital cortex was first discovered in the early 20th century (Holmes, 1918). Evidence for an orderly spatiotopic map in the human cortex for representing the visual field bore out of perimetry experiments using veterans of the First World War with head trauma sustained in combat. The orderly relationship between the location of scotomas and occipital lobe lesions led to the conclusion that the brain represents the visual field in terms of retinal coordinates. The archive of neurophysiological research owes much of its expanse to this early work. In the last 50 years, enormous strides

have been made in characterizing the representational topography of the visual world in the central nervous system.

The validity and efficacy of research exploring the neural basis of visual processing hinges on the capacity of the experimenter to faithfully survey the retinotopic functional organization throughout the cortical and subcortical pathways. For decades, the standard operating procedure for electrophysiology experiments investigating visual processing has been a two-phase enterprise. First, experimenters define the location, extent, and boundaries of a neuron's classical receptive field. Once the classical receptive field is experimentally defined, a given neuron's response properties are explored by comparing the relative neural activity when stimuli are presented inside versus outside the receptive field (Hartline, 1938; Hubel & Wiesel, 1963; Hubel & Wiesel, 1968). In the arena of functional brain imaging, this experimental approach can be adopted with relatively little modification. The real challenge is to adapt the classical receptive field model in such a way as to make it useful for brain imaging research, taking into account the disparity in both spatial and temporal scale between single-cell electrophysiology techniques and whole-brain magnetic resonance imaging.

The functional organization of early human visual cortex has been the topic of extensive neuroimaging research (Engel et al., 1994; Sereno et al., 1995; DeYoe et al., 1996; Tootell et al., 1997). FMRI offers a unique experimental avenue by which multiple visual areas can be functionally sampled concurrently,

allowing researchers to probe the response properties of multiple visual areas with a single dataset and analytic approach. Traditionally, using fMRI to map visual areas has relied on the traveling wave paradigm whereby periodic modulation in the visual stimulation accompanied by a frequency-domain analysis of the imaging data allows researchers to distinguish visually responsive voxels according to the frequency and phase of their responses. Presenting the visual system with a stimulus that varies periodically over time and space should entrain voxels to the periodicity of the stimulus. Since neurons in early visual areas have been shown to have retinotopically organized receptive fields, adjacent voxels—which presumably respond to adjacent portions of the visual field—show slight differences in the phase of their periodically entrained responses. Hence, transforming the time-series to the frequency domain via the Fourier transform and computing the phase at the particular stimulus frequency (say 8 cycles per scan) will inform the results about the spatiotopic relationship between the retina and cortex.

While the phase-encoding method for performing retinotopic mapping is robust and widely recognized within the field, it is limited in its precision for determining both the location and dispersion of the representation in the visual field. Since a the phase of a voxel reveals information about where in a periodic stimulus' cycle to which it's most responsive, two sets of runs are required for determining the polar angle and the eccentricity of the voxel. Later, that information can be combined to yield a screen coordinate for each voxel, but it is

impossible to determine to extent over which a voxel pools its response in the visual field. In order to measure and later utilize information about each voxel responsiveness in terms of the location in and the extent over the visual field, we invoked the population receptive field model (Dumoulin & Wandell, 2008; Amano et al., 2009).

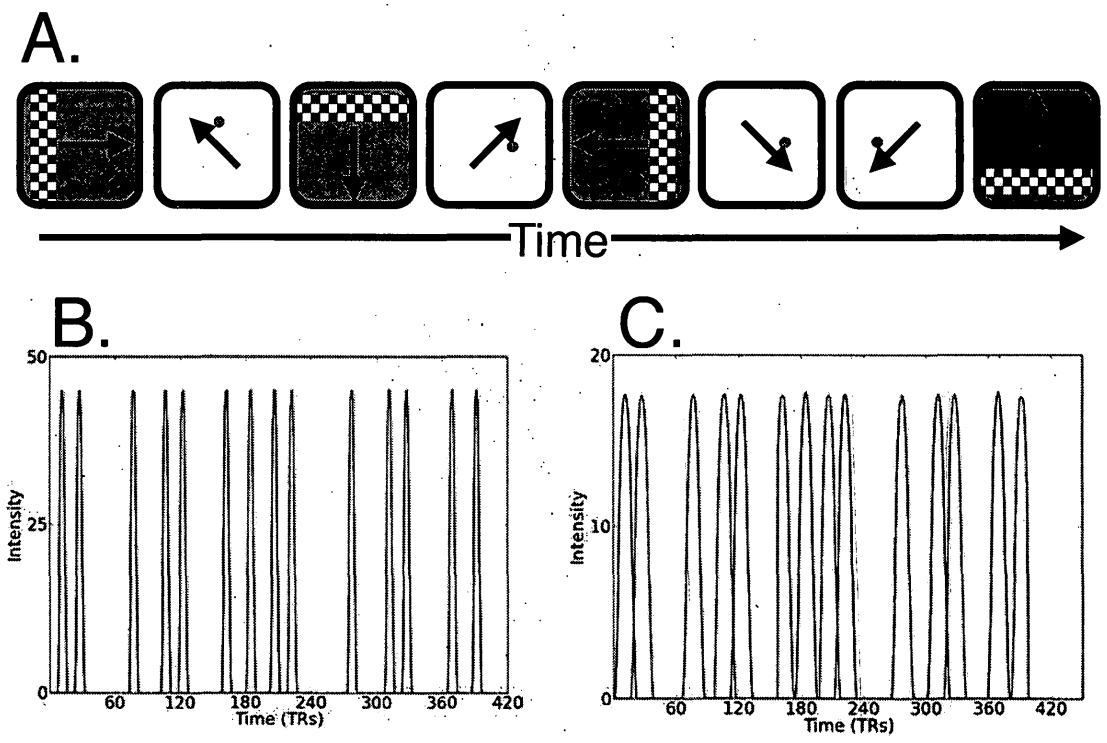

Figure 1. The sweeping bar stimulus transits the visual field in 8 different directions over the course of a single run (A). Assuming that each portion of the bar stimulates equally, differing responses can be elicited based on both the location and dispersion of a hypothetical population receptive field. Differentiable time-series are attainable by comparing the luminance intensity changes for two 1° patches of the visual field (B, red and blue). Changing the size of the patch to 5° (C) allows for the discrimination of two identical positions.

3.2 The population receptive field model

The population receptive field (pRF) model takes advantage of the fact that we know how a the response of a voxel should vary over a range of locations as well as receptive field sizes. The only requirements for such a methodological approach is that the visual stimulation varies over time and space and that it has a linear relationship with the measured neural activity. A toy example of how presumptive visual signals might vary across the visual is shown in Figure 1. Given a sweeping bar stimulus that systematically sweeps through the visual field, plotting the luminance changes that occur over the course of a run at locations (x_1, y_1) and (x_2, y_2) highlights differentiable luminance signals at each location. Furthermore, pooling the signals for multiple locations surrounding (x_1, y_1) and (x_2, y_2) yields an even subtler difference in the luminance signals.

This example offers an intuitive sense for how signals among voxels may vary according to the locations and dispersions of their responsiveness across the visual field. The pRF model is an encoding model in that the signal derived from each voxel is driven in a systematic way via the pattern of visual stimulation. The parameters of the model can be quantitatively worked out so long as the measured signals have a linear spatiotemporal relationship with respect to the stimulation (Boynton et al., 1996), and that the precise pattern of the stimulation is known during the analysis phase.

3.3 Population receptive field stimuli

The visual field mapping stimuli consisted of a moving bar aperture that revealed a high-contrast (100%) flickering checkerboard pattern, reversing contrast at a rate of 4 Hz. Over the course of a single run, the bar was presented at four orientations (0°, 45°, 90°, and 135°). The bar moved in the two opposite directions perpendicular to their orientation, yielding a total of 8 different bar sweeps per run (Figure 2). The projection system and bore sized allowed for 13° of visual angle from fixation. Each bar traversed 26° during a pass, always starting in the periphery, passing through fixation, and terminating in the periphery, The bar width was set to 1/4th of the presentation field radius (3.25°), as prescribed by previous pRF papers (Dumoulin & Wandell, 2008; Amano et al., 2009). The bar moved across the field in discrete 0.625° steps across the visual field, with each step being time locked to the collection of a new functional brain imaging volume. Each bar sweep took a total of 60 s to traverse the visual field.

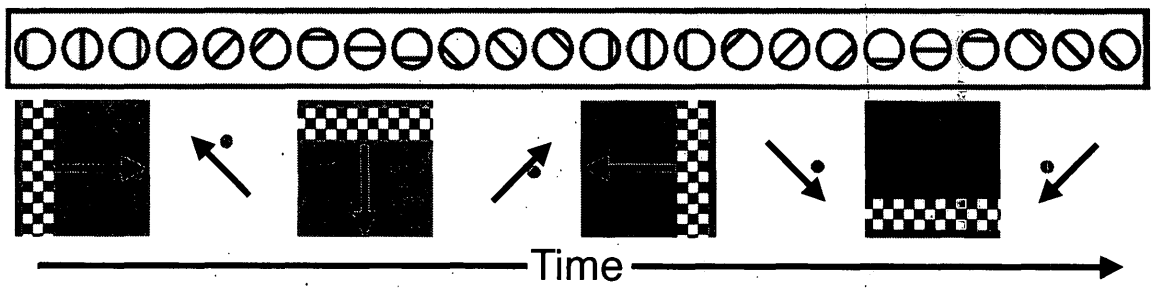

Figure 2. The visual field mapping stimuli consisted of a moving bar aperture that revealed a high-contrast (100%) flickering checkerboard pattern, reversing contrast at a rate of 4 Hz. Over the course of a single run, the bar was presented at four orientations (0°, 45°, 90°, and 135°). The bar moved in the two opposite directions perpendicular to their orientation, yielding a total of 8 different bar sweeps per run.

3.4 Population receptive field estimation

We modeled the fMRI signal as a function of both the location in the visual field to which a given voxel is responsive (x,y) as well as the dispersion of the response across the visual field σ. Hence, our 3-parameter model yields two-dimensional Gaussians which vary in location, extent, and amplitude as a function of dispersion across the visual field from voxel to voxel. The Gaussian function

$$g(x,y,\sigma) = exp-\left(\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}\right)$$

provides the formulation for the population receptive field where x_0 and y_0 represent the coordinate matrices defined by the horizontal and vertical dimensions of the visual display and σ denotes the standard deviation. The location and size of the Gaussian can be arbitrarily set by varying the parameters (x,y) and σ, respectively.

The effective stimulus (i.e., the sweeping bars) was characterized in terms of its position on the screen (x,y) at a discrete unit of time t. Assuming that all portions of the flickering checkerboard pattern contribute equally to the fMRI response (Engel et al., 1997), the predicted hemodynamic signal can be computed by multiplying the two-dimensional Gaussian by the stimulus frame at

a given time-step and summing the product. Doing so over all time-points yields an expected fMRI response (r) for a given model-parameter triplet (x,y,σ)

$$r(t) = \sum_{x,y} s(x,y,t)g(x,y,\sigma)$$

The expected response r was then convolved by the canonical hemodynamic response function (Glover, 1999; Friston et al., 1998; Worsley et al., 2002) to yield a predicted fMRI signal

$$p(t) = r(t) * h(t)$$

The goodness of fit was computed as the residual sum-of-squares (RSS) between the actual fMRI response and the prediction

$$RSS = \sum_{t} (y(t) - p(t))^2$$

In estimating the model parameters, two difficulties arose in turn: first, due to the multitude of possible solutions in estimating the model, an exhaustive brute-force search would be impractical; second, if a gradient-descent error minimization procedure for model estimation was employed, an initial starting point for the model estimation would have to be provided. I employed a two-phase, coarse-to-fine model estimation strategy that used an abbreviated brute-force search to discover the initial parameter settings for the gradient-descent error minimization

routine to converge on the final solution. All pRF estimation routines were recreated and reinvented by the author of this thesis. The software was written independent of the code base distributed by the originators of the pRF model and will be made available under a public use license.

The first phase of the parameter estimation approach involved a brute-force grid-search that regularly sampled candidate values in the parameter space. Because of the multitude of possible model solutions and the combinatorics entailed, the grid-search estimation procedure was handed a coarsely sampled effective stimulus and roamed a sparsely sampled parameter space. The effective stimulus was down-sampled to a resolution of 5% of the original using a three-dimensional linear interpolation. The procedure for sampling the parameter space was slightly more sophisticated. Here I implemented an adaptive brute-force search that would iteratively tighten the bounds of the three parameters via error minimization. Each parameter was given four evenly spaced dimension coordinates with which to generated a prediction. The RSS between it and the actual measured signal was computed for each of the 64 (4×4×4) permutations. The triplet with the lowest RSS was selected was selected as the seed for a new brute force within each of the three parameter dimensions, again using regularly spaced samplings along each dimension. At each iteration through this procedure, the distance between the samples became incrementally smaller at square rate. The initial distance between the bounds and the number of iterations could be set arbitrarily,

although certain search configurations may lead to an overwhelming penalty in terms of computational time. The point of the adaptive brute was to ball park an initial guess to use as a starting point for a gradient-descent error minimization.

The gradient-descent procedure used was a downhill simplex algorithm (Fletcher & Powell, 1963) for rapidly traversing down the error surface in the three-dimensional space of the model parameters. Once the solution was ascertained for a given voxel, the parameters were stored in conjunction with the covariance between the predicted and actual fMRI signal, for the purpose of thresholding.

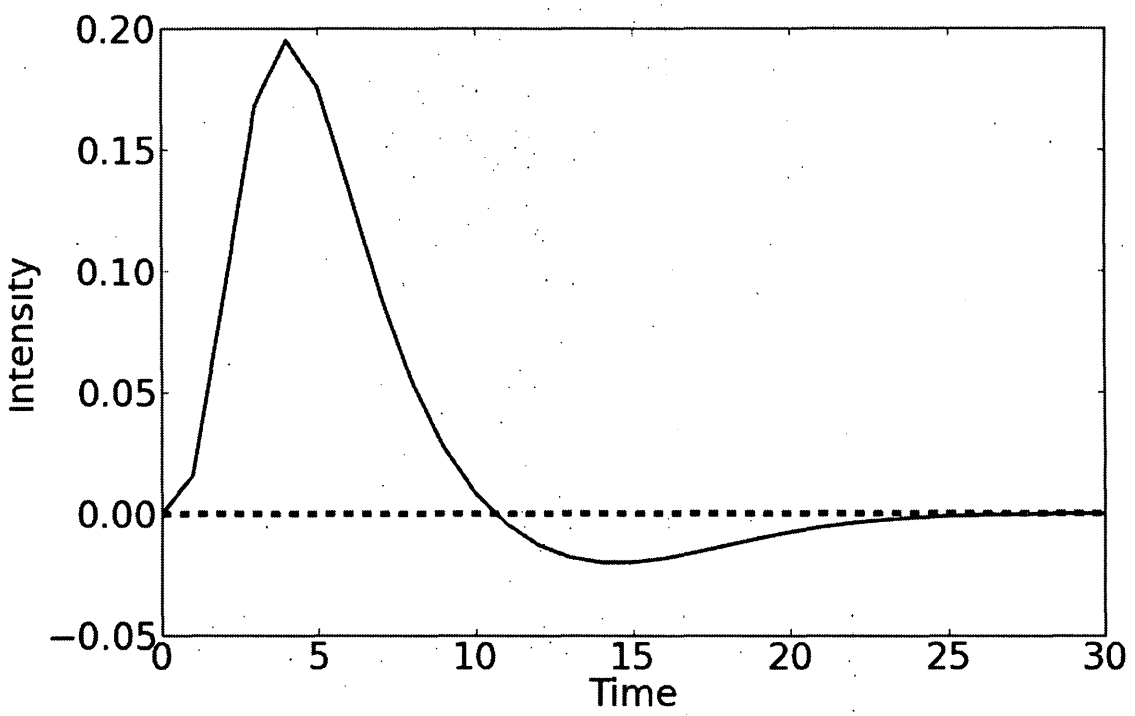

Figure 3. The canonical hemodynamic response function that is convolved with the stimulation time-series. In response to single brief stimulus, the function peaks 4-6 seconds post-stimulus. The signal then decays begins to undershoot the baseline (dashed line), reaching the most negative response about 14-16 seconds post-stimulus. At 22-24 seconds post-stimulus, the signal returns to baseline.

3.5 Voxel-wise HRF estimation

In addition to fitting the pRF model to the fMRI signal, we also incorporated an HRF estimation procedure for generating a unique HRF on a voxel by voxel basis. The HRF describes the evolution of the BOLD signal in response to some stimulus. While the formulations of the HRF vary greatly and a considerable amount of research has been devoted to determining the physiological basis and experimental consequences of a given HRF, versions of the HRF have emerged as standards in the field of fMRI and, so, have been dubbed "canonical". One of the more widely used HRFs is the double-gamma HRF, formulated as

$$h(t) = \left(\frac{t^{\alpha_1 - 1} \beta_1^{\alpha_1} e^{-\beta_1 t}}{\Gamma(\alpha_1)} - c \frac{t^{\alpha_2 - 1} \beta_2^{\alpha_2} e^{-\beta_2 t}}{\Gamma(\alpha_2)} \right)$$

where α_1 and β_1 represent the delay and dispersion of the peak, α_2 and β_2 represent the delay and dispersion of the undershoot, c represents the ratio of the peak to the undershoot, and Γ is a gamma function. Figure 3 shows the canonical HRF. In response to a hypothetical stimulus, the BOLD signal in some voxel will begin to increase. Approximately 4-6 seconds post-stimulus, the BOLD signal will peak and begin to decay. On the downslide towards the initial intensity, the BOLD signal will undershoot its starting point, and eventually return back to baseline.

In the coarse-search phase of the analysis, the HRF was assumed to be the canonical form. In fine-search phase of the analysis, however, the gradient-descent error minimization routine was allowed to explore a parameter space that included an HRF-delay term in addition to the three parameters of the pRF model. The delay term equally affected the peak and undershoot, effectively translating the HRF left or right along the ordinate (time) axis depending on whether the optimal delay was negative or positive. Since each stimulation run contained pairs of bar sweeps in mutually opposite directions (eight sweeps, four orientations), the delay parameter of the HRF model was not degenerative with the location estimate of the pRF model. If the stimulus was not designed with opposing pairs of bar sweeps, the fitting procedure would erroneously conflate the HRF delay parameters with the pRF location parameters.

3.6 Population receptive field results

The pRF model estimation routine generates a nine dimensional volume, comprised of the three spatial dimensions, the three pRF model parameter estimates, the HRF delay parameter, the residual sum of squared error between the actual time series and the model fit, and the covariance between the actual BOLD time-series and the model fit. The actual and predicted time-series are shown in Figure 4 for three example voxels from a particular participant. The voxels were chosen based on their anatomical locations and their location in the retinotopic space. Panels A and B show the actual and predicted time-series for

voxels selected from anterior and posterior portion of the calcarine sulcus. Previous experiments would suggest that these voxels contain neural representations of the periphery and fovea, respectively. The voxels differ from one another in a number of ways, including the distance between and amplitude of BOLD signal peaks. The model parameters denoted in each panel dutifully describe the subtle differences among the measured BOLD signal modulations of two voxels. The model parameter estimates confirm anterior-posterior foveal-periphery gradient retinotopic property of early visual cortex. The voxel in panel C was selected from dorsolateral occipital lobe and may perhaps be sampled from what is generally considered to be the motion-sensitive area, human MT+ (Wandell et al., 2007). This particular voxel is interesting because of the

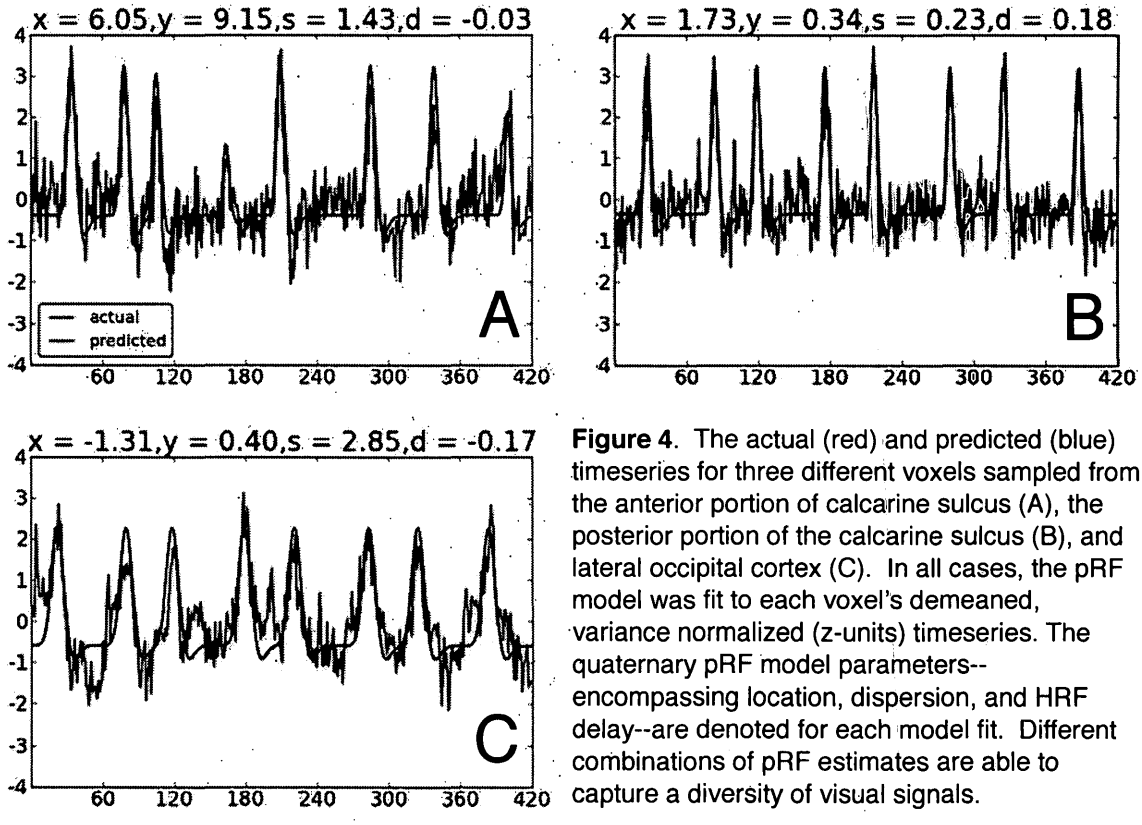

Figure 4. The actual (red) and predicted (blue) timeseries for three different voxels sampled from the anterior portion of calcarine sulcus (A), the posterior portion of the calcarine sulcus (B), and lateral occipital cortex (C). In all cases, the pRF model was fit to each voxel's demeaned, variance normalized (z-units) timeseries. The quaternary pRF model parameters--encompassing location, dispersion, and HRF delay--are denoted for each model fit. Different combinations of pRF estimates are able to capture a diversity of visual signals.

similarity between its location and HRF delay parameter estimates and those of the the voxel depicted in panel B. These voxels differ in their anatomical location in the brain and their receptive field size estimate. Retinotopically organized visual areas downstream of V1 were found to have increasingly larger and larger receptive fields (Harvey & Dumoulin, 2011). Cells in visually response cells in temporal and frontal cortices have been found to have receptive fields spanning entire an entire hemifield, and in some case, the totality of the visual field. Panel C demonstrates this effect in limited way, but also highlights the ability of the pRF model to map out the response properties of higher visual areas.

Figure 5 shows the location parameter estimates (x,y) of the four anatomical subdivisions of V1, including the left and right hemispheres and the dorsal and ventral calcarine sulcus plotted in the visual display space in terms of degrees of visual angle. The central scatter plot shows the locations of pRF estimates across all voxels in visual areas V1, V2, and V3 whose covariance between the actual and predicted time-series exceed 0.2. The histograms above and beside the scatter plot show the distribution of x and y, respectively.

Since the voxels were extracted from both hemispheres and the dorsally and ventrally to the calcarine, the location estimates are distributed across the entire visual field. The shapes of the distributions are roughly normal, although the distribution along the vertical dimension of the display space shows a moderate degree of kurtosis, with a Fisher kurtosis of 1.62 (Zwillinger & Kokoska, 2000).

The kurtosis suggests that visual cortex inequitably samples the visual space, with relatively few voxels representing the periphery along the vertical meridians. This sort of letter-box representation has been previously found in human brain imaging experiments across cortical and subcortical visual areas (Kastner et al., 2007; Schneider, 2011). Psychophysics experiments have also demonstrated a behavioral deficit in visual discrimination tasks in the periphery along the vertical meridian (Carrasco et al., 2001; Corbett & Carrasco, 2011).

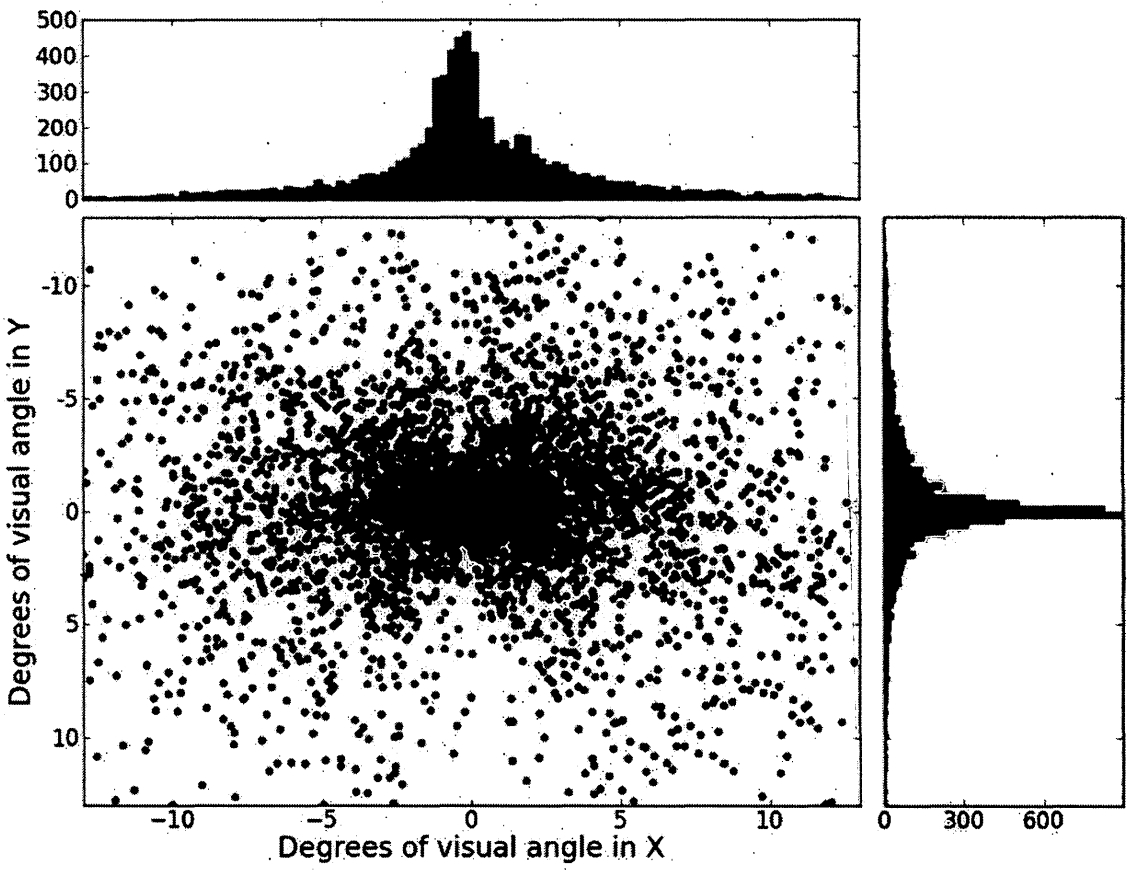

Figure 5. The distribution of location estimates across the entire visual space for participant P2. The central scatter plot shows the intersection of the x and y location estimates among voxels throughout the measured brain whose covariance between the actual and predicted timeseries exceeded 0.2. The distribution of the x and y estimates are shown in histograms above and beside the scatter.

While this view of the data provides a good idea of the global visuospatial properties of early visual cortex, it doesn't illustrate the functional topography of the cortex itself. To do this, it is necessary to project the model parameter estimates into the cortical space (Dale et al., 1999; Fischl et al., 1999). Figures 6 and 7 show the pRF estimates projected onto the cortical surface of a single participant, P7. Panels A, B, and C show the pRFs= projected separately onto the left and right hemisphere surfaces. For clarity, the cortical surface reconstructions from each hemisphere have been flattened and cropped so as to only include to posterior portion each hemisphere.

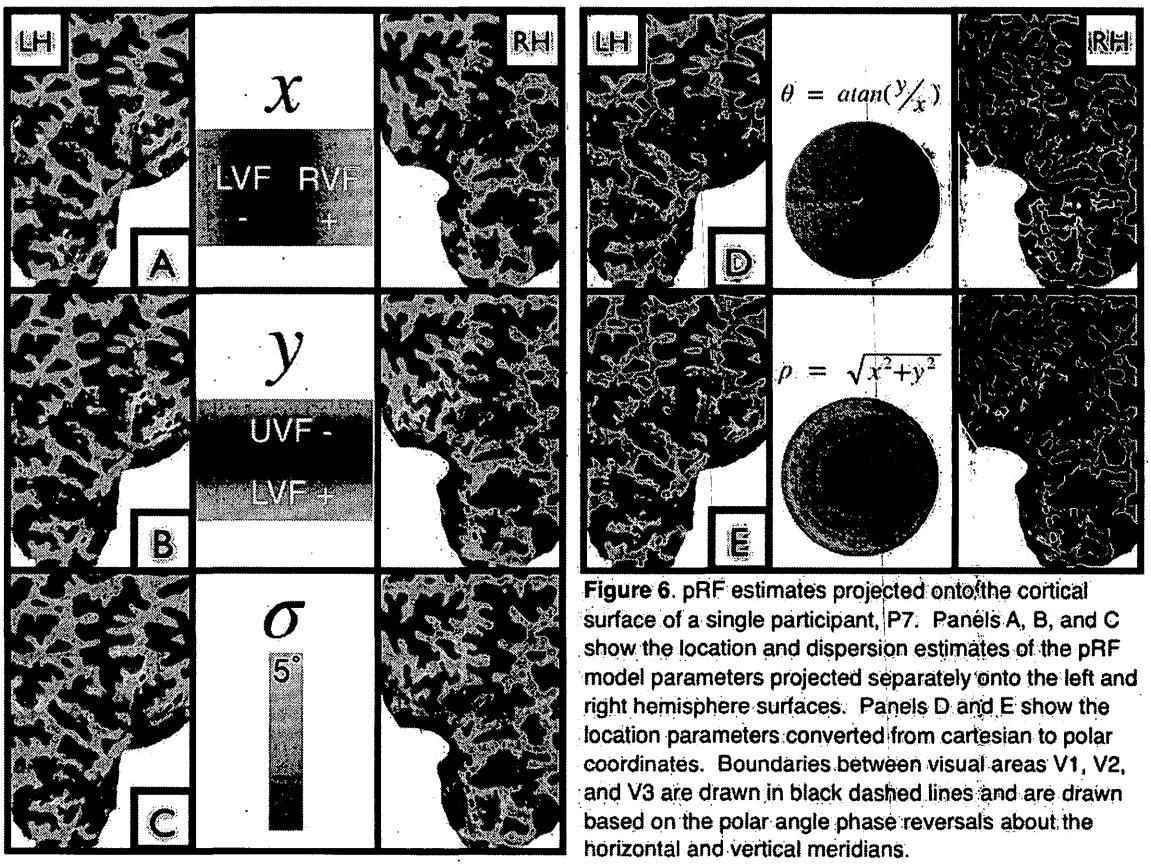

Figure 6. pRF estimates projected onto the cortical surface of a single participant, P7. Panels A, B, and C show the location and dispersion estimates of the pRF model parameters projected separately onto the left and right hemisphere surfaces. Panels D and E show the location parameters converted from cartesian to polar coordinates. Boundaries between visual areas V1, V2, and V3 are drawn in black dashed lines and are drawn based on the polar angle phase reversals about the horizontal and vertical meridians.

The values of the location estimates (x, y) vary about 0, with positive and negative values traversing fixation horizontally and vertically. Panel A reaffirms the basic retinotopic property of contralateral visual field representation in early occipital cortex; hence, the horizontal estimates project onto the right hemisphere range from fixation to the extreme periphery of the left visual field. Estimates ranging from fixation to the extreme periphery of the right visual field project onto the left hemisphere cortical surface. Panel B shows the cortical projection of the vertical position estimates using the same colormap rotate 90˚. The results reconfirm the results of numerous experiments mapping the retinotopic organization of human and macaque cortex, with the upper and lower visual

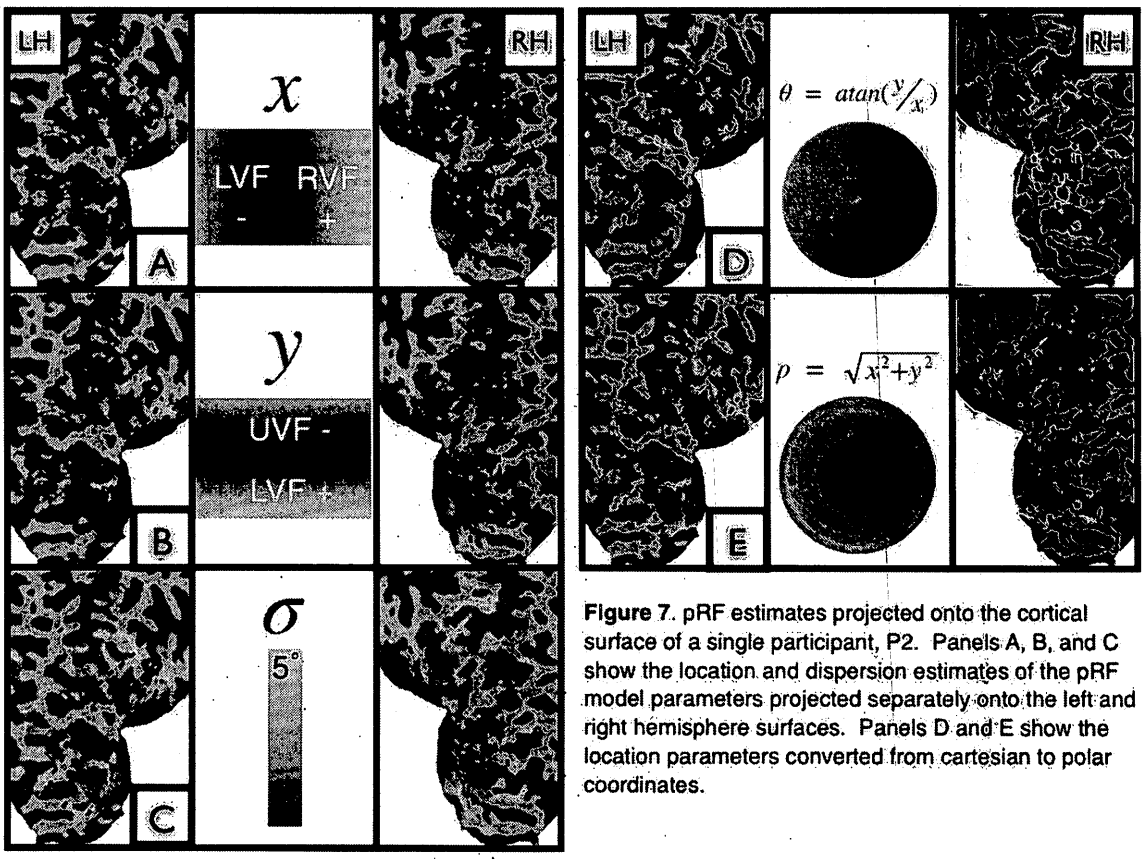

Figure 7. pRF estimates projected onto the cortical surface of a single participant, P2. Panels A, B, and C show the location and dispersion estimates of the pRF model parameters projected separately onto the left and right hemisphere surfaces. Panels D and E show the location parameters converted from cartesian to polar coordinates.

fields projecting to cortical zones ventral and dorsal to the calcarine sulcus. Panel C shows the pRF size estimates. Here, values were always greater than 0°, with the colormap being truncated at 5°.

Panels D and E show the location estimates location estimates in panels A and B converted from cartesian to polar coordinates. The polar angle map—measured in terms of degrees or radians—is shown in panel D. The polar angle be computed from the cartesian coordinates as $\theta = atan(y/x)$. Polar angle complements eccentricity, which describes the central to peripheral dimension of polar space. Eccentricity is computed as $\rho = \sqrt{x^2 + y^2}$.

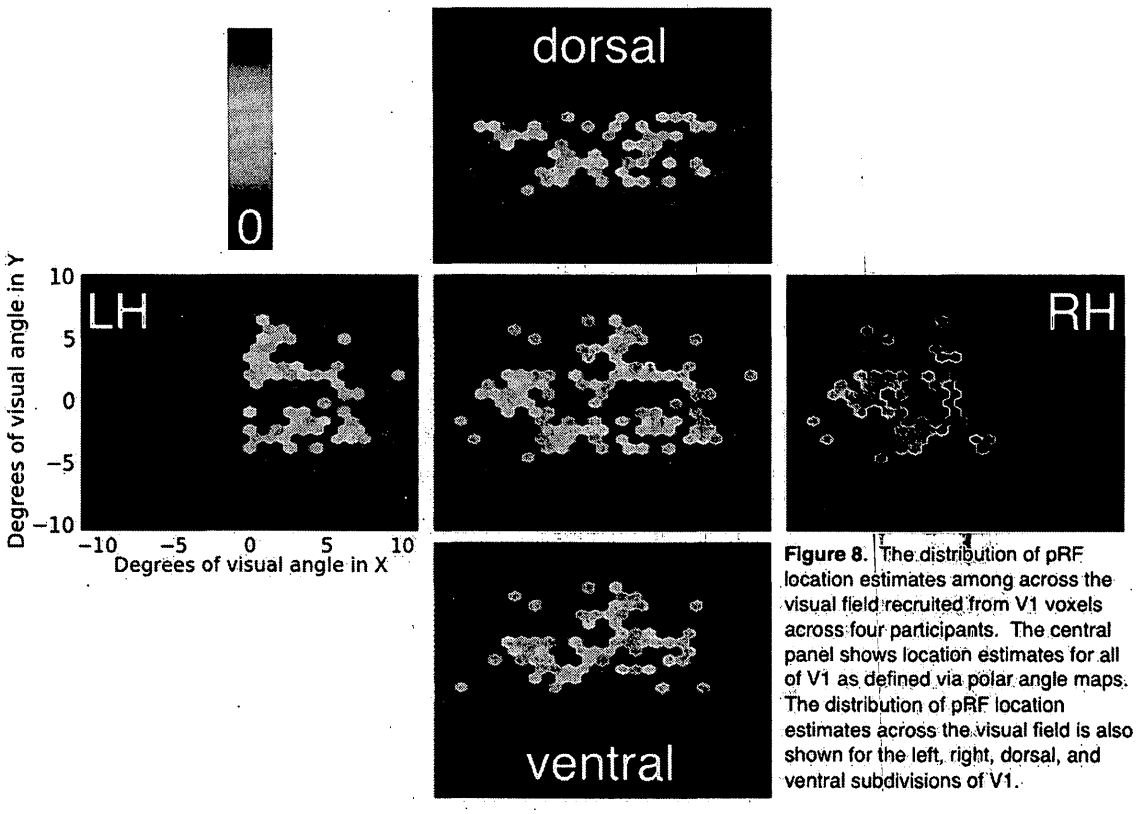

Figure 8. The distribution of pRF location estimates among across the visual field recruited from V1 voxels across four participants. The central panel shows location estimates for all of V1 as defined via polar angle maps. The distribution of pRF location estimates across the visual field is also shown for the left, right, dorsal, and ventral subdivisions of V1.

The organization of retinotopic cortex is notoriously unwieldy among the healthy control population, with the topography varying across individuals in terms of the functional size, location, and orientation of visual areas with respect to anatomy. These irregularities among areas and brains has plagued various attempts to automate the retinotopic parcellation of visual cortex via software solutions, both in three-dimensional volume and two-dimensional cortical surface space. Manual tracing remains as the state of the art in retinotopic cortical parcellation.

Boundaries between visual areas V1, V2, and V3 are drawn in black dashed lines. The boundaries are drawn based on the polar angle phase reversals about the horizontal and vertical meridians. Each of these visual areas

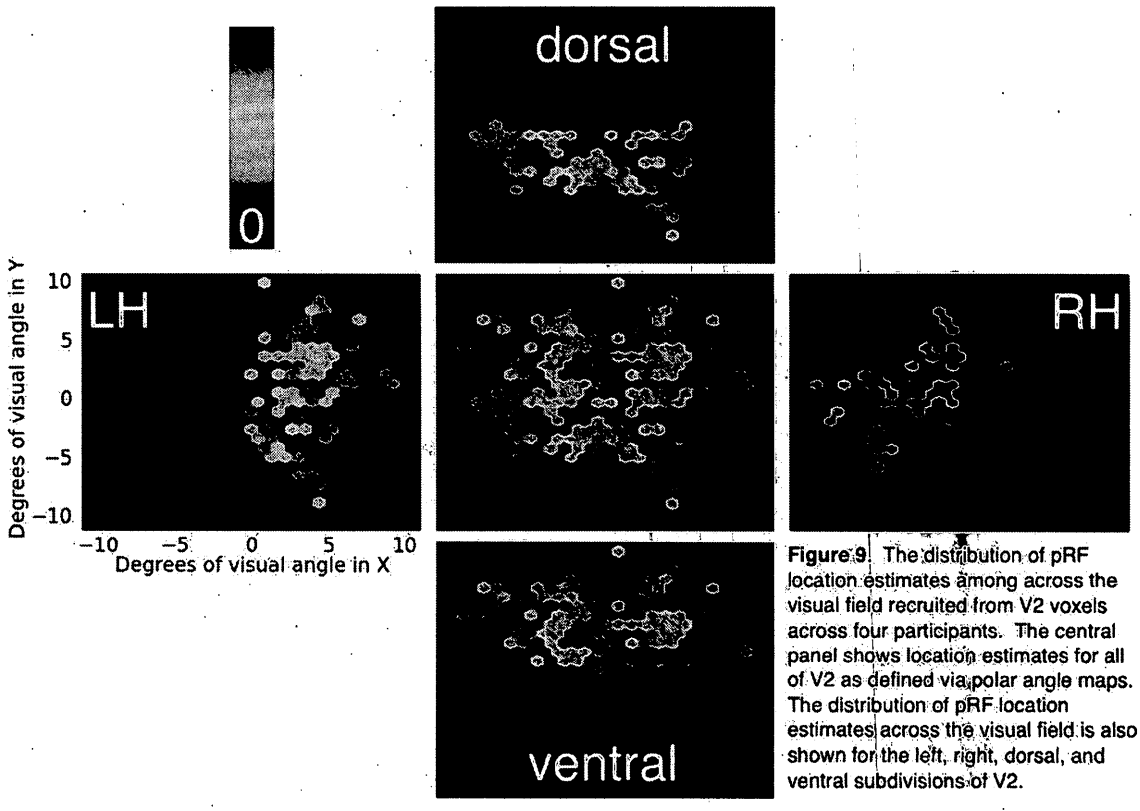

Figure 9. The distribution of pRF location estimates among across the visual field recruited from V2 voxels across four participants. The central panel shows location estimates for all of V2 as defined via polar angle maps. The distribution of pRF location estimates across the visual field is also shown for the left, right, dorsal, and ventral subdivisions of V2.

can be divided into quadrant representations of the visual field according to hemisphere and ventral-dorsal position, containing representations from the vertical to horizontal meridian along all eccentricities. Figures 8, 9, and 10 show the distribution of pRF location estimates around the visual field for the four subdivisions of each visual area combined across four participants (P2, P7, P23, P27). These figures illustrate the relationship between the pRF location estimates among each of visual areas' subdivisions and the visual field.

In addition to this first order property of retinotopic organization in the cortex, visual cortex also demonstrates a collinearity between the eccentricity of the visual representation and the receptive field size. Figure 11 shows this collinearity among multiple visuals are four participants. Here, each visual area

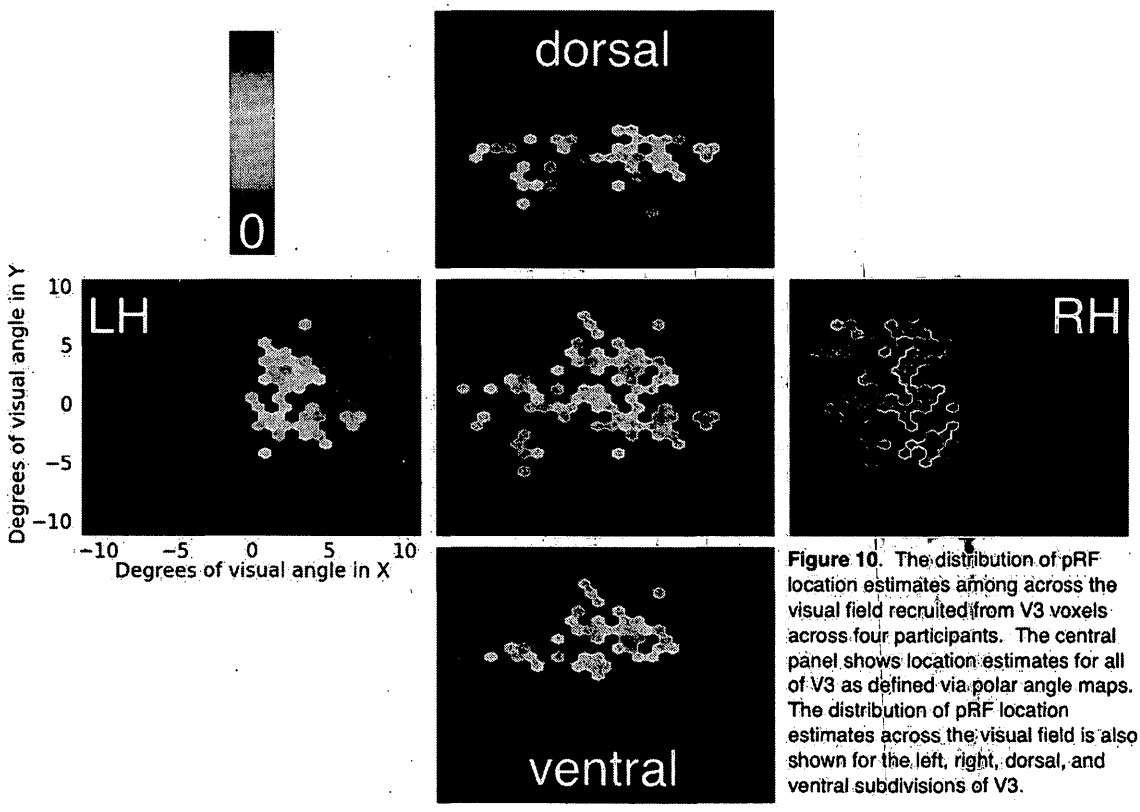

Figure 10. The distribution of pRF location estimates among across the visual field recruited from V3 voxels across four participants. The central panel shows location estimates for all of V3 as defined via polar angle maps. The distribution of pRF location estimates across the visual field is also shown for the left, right, dorsal, and ventral subdivisions of V3.

was taken to comprise both hemispheres and the dorsal and lateral subdivisions. Moving from foveal to peripheral representations among the visual areas, there is a corresponding increase in the receptive field sizes, denoted by the positive slope of each line. The intercept of this relationship differs among the visual areas, with smaller receptive field sizes at a given eccentricity for V1 compared to V2 and V2 compared to V3. This relation between the eccentricity and receptive field size maps—both within and across visual areas—has been

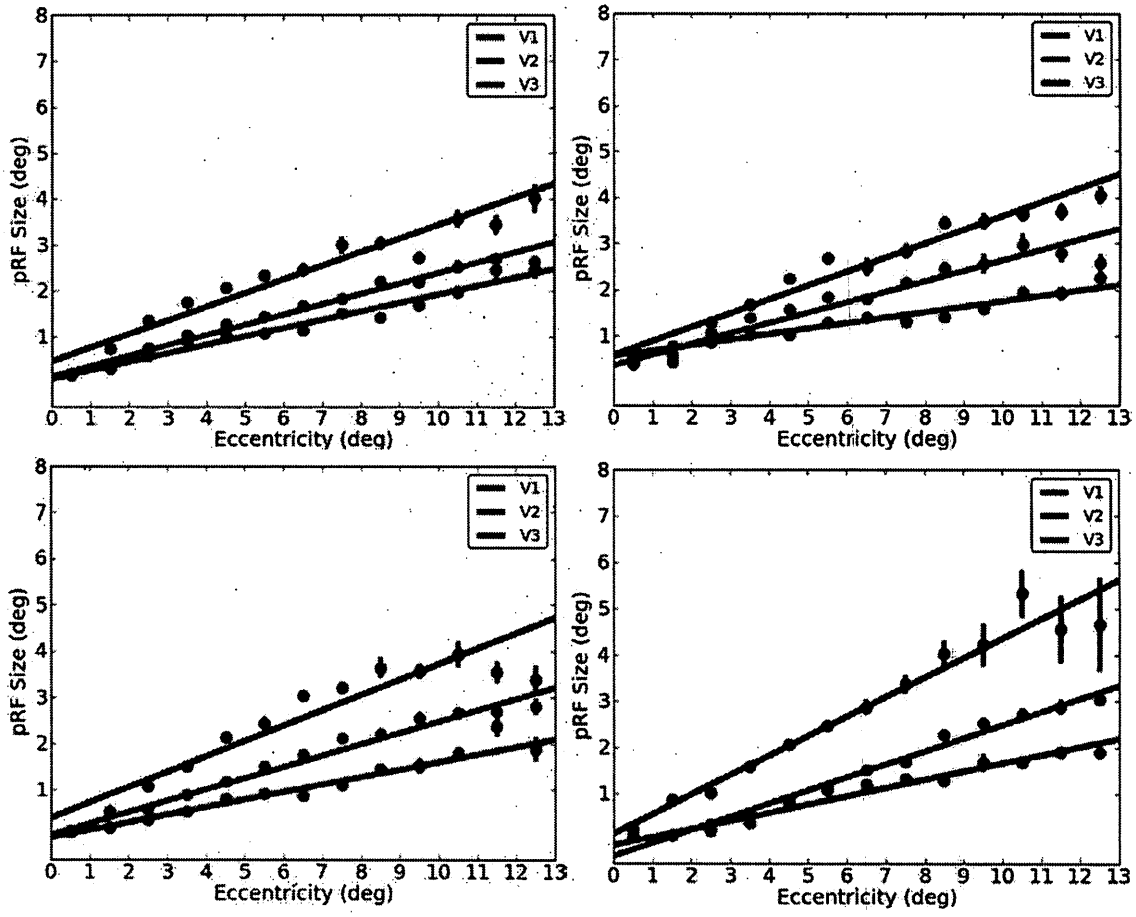

Figure 11. The relationship between eccentricity and receptive field size among multiple visual areas in four participants. Moving from foveal to periphery, there is a tonic increase in the pRF size, denoted by the positive slope of each line. The intercept of each line differs among the visual areas, with increasing pRF size at a given eccentricity along the V1, V2, V3 trajectory.

previously demonstrated in macaques using electrophysiology techniques and in humans using functional brain imaging (Amano et al., 2009; Harvey & Dumoulin, 2011).

3.7 HRF results

In addition to modeling the location and size of the pRFs, we explicitly modeled the delay of the hemodynamic response function. We allowed the gradient-descent error minimization to explore delay as a model parameter. As mentioned earlier, since the bar sweeps occurred in opposing pairs, the location parameters are not degenerate with the delay parameter. The distribution of HRF delays for four participants are shown in FIgure 13. The HRF delay values are negative or positive with respect to the baseline 5 s delay that is typically used as a good good approximation across the brain and the population. Kernel density estimation was used to estimate the continuous distribution of pRF delay estimates. Here, a kernel bandwidth of 0.25 was used. The delay distributions across visual areas are roughly normal and centered about 0 s delay. There is a slight shift from positive to negative delays and broadening of the dispersion of delays moving from V1 to V2 to V3.

3.8 Retinotopic mapping discussion

The results of the pRF modeling of visual cortex across multiple participants indicates that the code base of the original authors have been faithfully and

independently replicated. We were able to model the response properties of voxels sampled from various location in the brain (Figure 4). In addition, we able to define the boundaries of multiple visual areas across occipital lobe Figures 6 and 7) and that these cortical visual areas appear to show the expected registry with the visual space (Figures 8, 9, and 10). The second order relationship between eccentricity and receptive field size was also replicated among multiple visual areas and participants (FIgure 11). Interestingly, we were able to find an

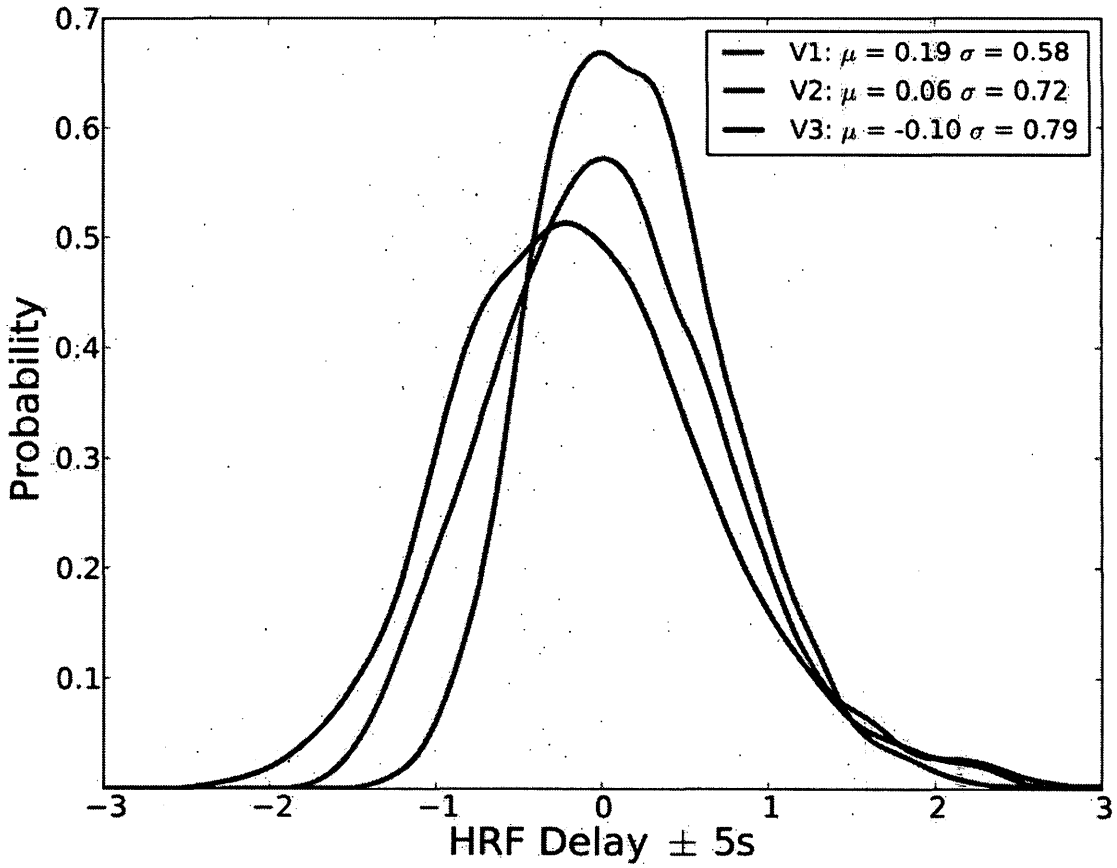

Figure 12. The kernel density estimation for the HRF delay estimates among visual areas V1, V3, and V3. The kernel bandwidth used was 0.25. The abscissa denotes the probability at any given delay estimate. The HRF delay mean and standard deviation for each visual are is shown in the legend.

orderly distribution of HRF delay estimates across voxels in areas V1, V2, and V3 (Figure 12).

These results are largely confirmatory in and of themselves. However, in the current context, the pRF maps are intended to serve as stepping stones for investigating the neural representations emanating from a tracking task. In order to do so, we need to recast the pRF maps in terms of the visual space—not the cortical space. This will allow us to examine how different portions of the visual field may be over- or under-represented in response to targets, distractors, occlusion, and nonocclusion. To this end, we developed a neural decoding method for generating stimulus reconstructions based on the pRF maps and the neural activity.

4. Neural decoding

Stimulus reconstruction falls under the broad rubric of neural decoding. It is a method for determining the relationship between known sensory inputs and measured neural responses such that one can be derived from the other. Stimulus reconstruction is simply a flavor of neural decoding, and it is useful to highlight the distinctions between it and several other flavors of neural decoding approaches currently used in functional brain imaging research.

4.1 Image classification

In a seminal paper in the field of brain imaging, Haxby and others (2001) developed an analytic framework for mapping the patterns of voxel activity to categories of visual stimuli. Here, the researchers were interested in determining whether object categories (e.g., faces, houses, tools) could be distinguished from one another based on the distributed and overlapping pattern of voxel activity across brain networks spanning the lateral occipital and inferior temporal cortices. The participants were presented with images of exemplars from multiple object categories. A standard general linear model was used to regress out the effect among all voxels of each of the category types. The participants were then exposed to a new round of exemplars. The experimenters used the pattern of activity associated with each object category derived from the training dataset to determine which stimuli caused a given pattern of brain activity in the test dataset. This procedure worked incredibly well, accurately predicting when a

given brain activity pattern resulted from viewing houses and faces at a rate of 100%. The lowest accuracy observed for identifying exemplars categories based on brain activity was 65%, well above the 50% chance guess rate. Furthermore, the experimenters found that networks that responded maximally to a particular category type could be used to accurately predict the presentation of exemplars belonging to other, non-selective categories. This result points towards the importance of distributed networks—rather than specialized modules—in representing a multitude of features associated with a high-level visual stimulus.

This approach is broadly known as image classification and relies on the method of multi-voxel pattern analysis. In general, the experiments and analyses are designed such that classes of stimuli can be characterized and differentiated in terms of configurations of voxel activity patterns. The paradigm has been extended to other domains in visual neuroscience. Kamitani and Tong (2005) used a multi-voxel pattern analysis technique to train a classification algorithm to identify sets of voxels whose responses were found to code the specific orientation of gratings, an elusive finding in brain imaging. Kamitani and Tong reasoned that, although neurons specifically selective for certain orientations may have a functional topography too fine for typical functional imaging resolution to register, randomly sampling from a portion of visual cortex may result in a preponderance of neurons selective for a specific orientation to reside in a given voxel by chance. Given this unequal distribution, voxels may differ from one another according to their weak but predominant selective responses to

various values of some feature dimension, for instance sinusoidal gratings oriented at a particular angle.

Their voxel classification algorithm was trained using functional datasets collected while participants viewed 45° and 135° oriented sinusoidal gratings. On subsequent functional runs, the two orthogonal gratings were combined into a single plaid and presented to participants who were cued to attend to either the 45° or 135° grating component of the plaid stimulus. Kamitani and Tong found that their classifier was able to accurately predict which of the two plaid components the participants were attending during any given epoch based on the patterned time-course activity of over 800 voxels sampled from visual areas V1, V2, V3, and V4.

While the mind-reading aspect of these kinds of experiments is quite appealing, the important findings are related to the neural decoding that mediates the sensory input and the neural activation patterns. The classification algorithms are indeed able to learn the association between types of sensory inputs and patterns of voxel activity; however, the question becomes: does this tell us anything about the kinds of information being represented differently for faces versus houses, or for lines oriented one way versus the other? Where this approach falls a little flat is in its inability to classify *new, untrained* stimuli. Since the relationship between the training and test datasets is purely statistical, it becomes difficult to determine which features of a particular stimulus are being represented by a given pattern. Indeed, the only conclusion that can be drawn is

that two patterns of activation are distinct from one another and that they are each associated with a particular kind of stimulus. The details of how a pattern of activation is associated with a given stimulus is not explicitly modeled based on the stimulus features themselves, and so cannot be unambiguously determined.

4.2 Image identification

Image identification is a flavor of neural decoding that explicitly models the response properties of voxels. Kay and others (2008) used image identification to predict which particular image in a set of novel, natural scenes a participant viewed. Their approach consisted of two phases. In the first phase, each participant viewed a large training dataset consisting of 1,750 natural images. The activity of each voxel in response to these images was modeled based on a Gabor wavelet pyramid, describing the tuning along the dimensions of space (x,y,σ), orientation, and spatial frequency. This model is more comprehensive than the pRF model, which only describes the three spatial parameters (x,y,σ).

Once the model parameter estimates were derived for each voxel, the participants were submitted to a testing phase. Here, participants viewed 120 novel natural images similar but not identical to the training images. The set of model parameter estimates for a given cluster of voxels were then used to predict the aggregate voxel pattern activity evoked by each of the 120 novel images. Images were identified as having been viewed by the participant when

the correlation between the predicted activity and the actual activity was highest. Image identification performance peaked for one particular participant at 92% accuracy; however, this number diminished greatly when images were identified using the correlation between the predicted response and a single-trial response (32% accuracy), but was still performed at a level vastly superior to chance (<1%).

The distinguishing feature of image identification is that the predicted pattern of activity is based on models that explicitly capture the response properties of voxels. Whereas image classification attempts to associate patterns of activity with particular sets of stimuli, image identification uses the model parameters for various dimensions (e.g., space, orientation, spatial frequency) to generate a predicted pattern of activity. Novel images can be processed through the model parameter estimates of voxels to yield a predicted signal.

4.3 Image reconstruction

An interesting consequence of using the pRF model is the ability to reconstruct visual stimuli based on the pRF estimates and the time series of voxels. The pRF model deploys an encoding procedure that uses a three parameter Gaussian to describe the response properties of voxels whose signals are driven by known patterns of visual stimulation. Each voxel views the world through a uniquely situated aperture in the retinotopic space. pRF estimates are

computed for each voxel, describing the location and extent of the aperture through which the voxel reads in the visual space. The model can be easily inverted such that a given pRF estimate and signal intensity fluctuations can be combined to predict the luminance changes in the display space at specific locations and times. Such a procedure, when summated across a multitude of voxels for each step in the time series, generates a prediction of the visual stimulus presented to the participant and encoded in the intensity modulations of the voxels. This sort of stimulus reconstruction can be used on any novel dataset, the only prerequisite being that pRF estimates must have been already computed for a given participant. Furthermore, since groups of voxels can be functionally bundled into visual areas according to an arbitrary ruleset, stimulus reconstructions may be generated that highlight differences in the components of a visual task that various visual areas extract and represent.

Using the mapping between sensory input and brain measurements to generate a reconstruction of an organism's sensory processing isn't new. One of the earliest demonstrations of stimulus reconstruction within the arena of neuroscience was published more than two decade ago (Bialek et al., 1991). Recording from a motion-sensitive cell in the blowfly, the researchers were able to reconstruct the analog temporal profile of the stimulus (i.e., the velocity a moving stimulus) using the digital spiking output of single neuron. The reconstruction algorithm assumes a specific mapping between the digital spiking output of the neuron and the graded intensity modulation of the stimulus (in this

case in the domain of motion), of which rate coding is only one possible mapping. Other solutions exist for relating the spiking output of a neuron to the stimulus time series that generated it, most of which center around the likelihood of different spiking outputs given the stimulus. However, while the approach taken by Bialek and others was foundational for developing analytic linkages between neural signals and sensory inputs, its purview is quite limited. The kinds of stimulus reconstructions I'm interested in developing for use with functional brain imaging datasets should be generative for more complex spatiotemporal stimuli and for larger ensembles of neurons.

Stanley and others (1999) developed a stimulus reconstruction procedure for rendering natural images (e.g., faces, wooded scenes) from the spiking output of multiple neurons in the lateral geniculate nucleus (LGN) of the thalamus of an anesthetized cat. Such an approach is appealing in two ways: first, in order to extract meaningful information about the wide-ranging spatial and temporal properties of typical visual stimuli from brain measurements, a stimulus reconstruction approach that considers circuits over neurons should be requisite; second, such reconstruction algorithms should at least be extensible to decode brain activity for natural stimuli. The experimenters recorded from 7 LGN neurons simultaneously using a multi-electrode array. The experiment was repeated numerous times whereby video clips of white noise, a human face, or a wooded scene were presented. Cells were screened on the basis of the reproducibility of the measured neural response over repeated video clip

presentations. After the cull, experimenters were left with 177 cells that were used to derive linear filters for relating the spiking output of sets of neurons to the luminance changes of sets of pixels. Once this mapping was determined, stimulus reconstruction simply became a matter of modulating the luminance of the appropriate set of pixels for a given cell based on its spiking output. The reconstructed stimulus was generated by convolving the spiking output of a set of cells with their corresponding linear filters and summating over the set. The filters were optimized so as to minimize a cost function (i.e., the residual sum of squares) relating the actual stimulus and the reconstructed stimulus. For each pixel in the reconstructed stimulus, an average of 14 cells' responses were used. The neurons used to create the reconstructions included those with both on- and off-center receptive fields. In the end, the experimenters were able to generate stimulus reconstructions whose spatial correlation coefficients with the actual stimuli were as high as 0.8.

This experiment, and in particular the stimulus reconstruction procedure used, is interesting in the current context because of the similarity between the pRF model and the response properties of LGN cells. The experimenters mapped out the location and extent (x,y,σ) of the receptive field of a given neuron, including both the center and the surround. These parameter estimates were used to make an assumption about the number of neurons representing a particular position (i.e., screen pixel) in the visual space. Their implementation of

the image reconstruction required two components: 1) the neural responses of sets of LGN cells in response to visual stimulation; and 2) the mapping between neurons and screen coordinates. The thing they got for free by dint of using LGN cells was the receptive field structure and response properties of the cells themselves. As on-center cells became more active, the experimenters presumed an increase in the luminance for pixels which fall in the center of a given cell's receptive field and presumed a decrease in the luminance for pixels which fall in the surround. If one repeats this procedure over a multitude of neurons and summates the result, the product is a three dimensional dataset that resembles—both temporally and spatially—the stimulus used.

Such an approach should work for the voxel-based pRF model as well, albeit with somewhat different assumptions. In both the cases of individual neurons recorded via electrophysiology and populations of neurons recorded via fMRI, each observational unit is described in terms of a receptive field location and size. Once derived, the receptive field estimates across all voxels could be used to generate a stimulus reconstruction through the summation of intensity-weighted Gaussians across space in a manner nearly identical to that described by Stanley and others' 1999 study.

4.4 Stimulus reconstruction method

The lynchpin of the stimulus reconstruction procedure is the voxel-wise pRF estimation computed per participant using the sweeping bar aperture stimulus.

Each participant's pRF estimation volume is 7-dimensional, whereby each voxel contains seven different values. The first three dimensions denote the spatial location of a particular voxel with respect to the coordinate system of the MRI magnet bore. The remaining four dimensions denote the pRF model parameter estimates for the location and extent of the receptive field (x,y,σ) as well as the delay estimate d of the double-gamma HRF. These four estimates describe not only *where* a given voxel is responsive to in the visual field but also *when* a given voxel is responsive in time following a stimulation event.

To begin, it is helpful to review the nature of the actual stimulus as it was drawn to the screen and stored for later use in the pRF estimation. The actual stimulus was displayed on a screen and projected at a resolution 800×600 pixels. While the refresh rate of the display was set to 60Hz, the sweeping bar stimulus was designed in such a way as to be phase-locked to the repetition time of the scanner (i.e., 1500 ms). As each new volume was acquired every 1500 ms during the course of a functional run, the stimulus would advance by discrete intervals at the outset of every new volume acquisition. The stimulus, then, could be captured in terms of screen coordinates and binned into time-units equal to the duration of the functional volumes. Hence, while our functional runs were of size 128 ×128×22×440 our stimulus runs were of size 800×600×440. The parity in the time dimension between the functional and stimulus datasets allowed us to relate the neural activity to the stimulation pattern via the pRF

model. This is crucial not only for determining the response properties of voxels, but also for recreating the stimulus based on the pRF estimates and signal intensity modulations of the voxels.

Once the pRF model estimates had been computed for every voxel, we set about predicting the stimulus shown to the participant from the signal intensity modulations of the voxels via image reconstruction. The first step in the image reconstruction process was to initialize an empty matrix with the same dimensions as the actual stimulus (i.e., 800x600x440). The predicted stimulus as this point is essentially a representation of the participant viewing a 800x600 display with zero mean luminance. For each acquisition in the functional run, we iterated through each of its constituent voxels. A particular pRF estimate

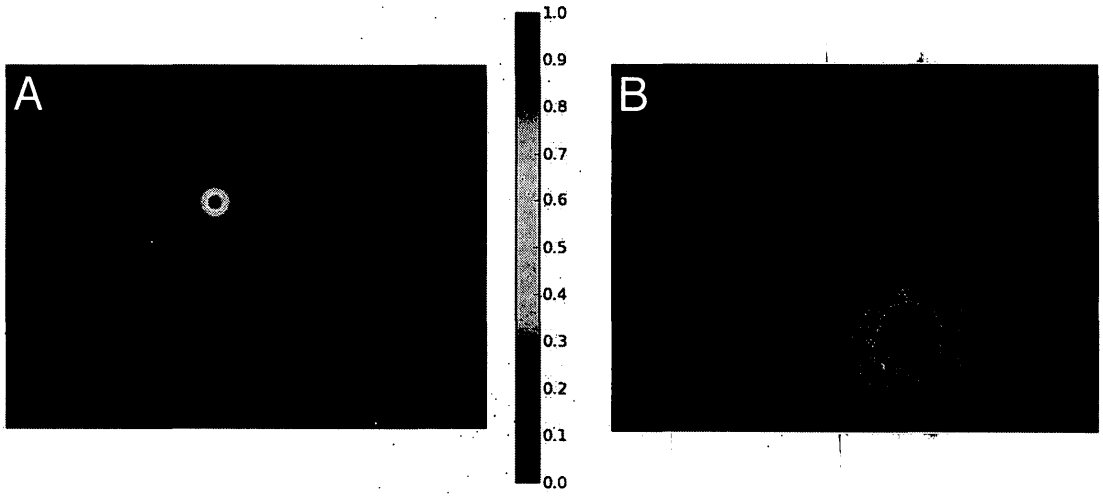

Figure 13. The pRF estimates of two voxels plotted in screen coordinates. The pRF in panel A was generated with parameters (-2.4, -3.2, 0.7). The pRF in panel B was generated with parameters (4.2, 6.3, 3.7). The colorbar shows the units of each of the Gaussians generated from the pRF estimates. In both panels, the red dot in the center represents the fixation point.

provides enough information to recreate a single Gaussian at a particular location and of a particular size among the screen coordinates. Figure 13 shows the pRF estimates for two different voxels plotted in the screen coordinates that match the stimulus dimensions from which the estimates were derived. Notice how the location and extent parameters affect the Gaussians that are computed from the two pRF estimates. Natively, the units of the screen coordinates were pixels. For the purposes of deriving the pRF estimates and stimulus reconstruction, the screen coordinates were rescaled to degrees of visual angle. Therefore, pRF estimates whose x-coordinate is negative are plotted to the left of fixation and whose y-coordinate is negative are plotted above fixation. The fixation point is also drawn in both pRF plots.

In order to reconstruct a single frame of our stimulus, we iterated through all relevant voxels and generate the Gaussian from each of their pRF estimates. We used the signal intensity of the voxels at any given point in time as a graded indication of the level of activation of that voxel and, hence, of the luminosity of the stimulus in the corresponding portion of the visual field. The equation for computing the scaled Gaussian then becomes

$$g(x,y,\sigma,i) \quad = \quad exp - \left(\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma^2} \right) i_t$$

where i_t represents the voxel signal intensity i at time t. Since the distribution of voxels' signal intensities at any given point in time will range in both positive and

negative values, it is possible to generate Gaussian distributions above and below zero, depending on the signal intensity by which they are scaled. Figure 14 shows how the Gaussians generated from two pRF estimates interact when combined into a single screen frame of the stimulus reconstruction. Panels A and B show the Gaussians produced when two pRF estimates are rendered. Although the pRF estimates are on either side of the fixation point with locations centered on (-3˚,2˚) and (3˚,2˚) respectively, they overlap slightly because of the

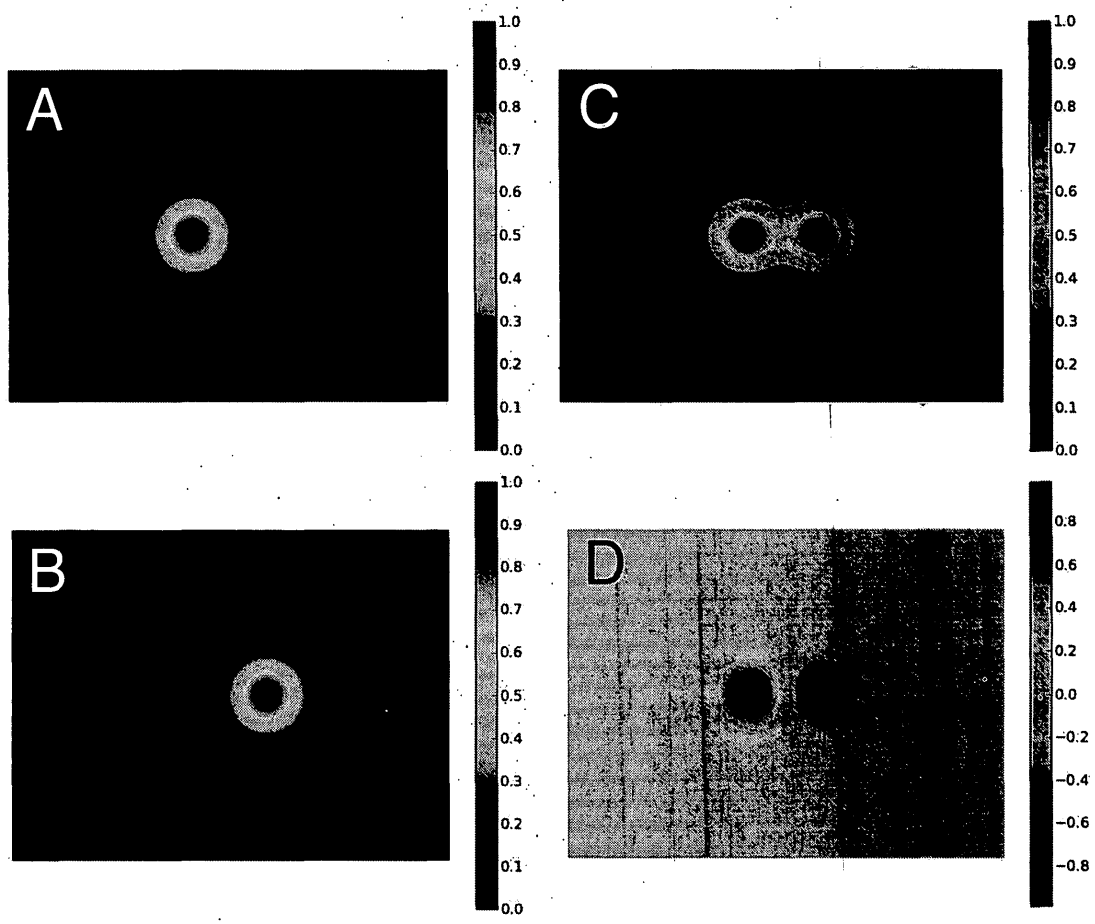

Figure 14. The pRF estimates plotted in the visual space. Panel A shows the Gaussian generated from a pRF estimate of (-3,0,2). Panel B shows the Gaussian generated from a pRF estimate of (3,0,2). Panel C shows the summation of the Gaussians shown in panels A and B when scaled by +1. Panel D shows the result of adding the two Gaussians from panels A and B when scaled by +1 and -1, respectively.

extent of their Gaussians (2°). If each pRF estimate is scaled by 1 and summated, the result is an bimodal distribution as shown in panel C of Figure 14. More interesting summations can be produced by scaling the Gaussians with differently signed intensities as shown in panel D. Here, the result is again a bimodal distribution; however, scaling the Gaussians with a negative and a positive intensity and summating across the two results in a distribution of pixel values for this particular stimulus reconstruction frame that spans both positive and negative values. Additionally, using both negative and positive scaling factors and summating the results will begin to produce stimulus reconstructions with sharper edges than any of their constituent Gaussians possess. Figure 15 shows how different configurations of Gaussians scaled by negative and positive intensities can produce interesting reconstructions with rich spatial information.

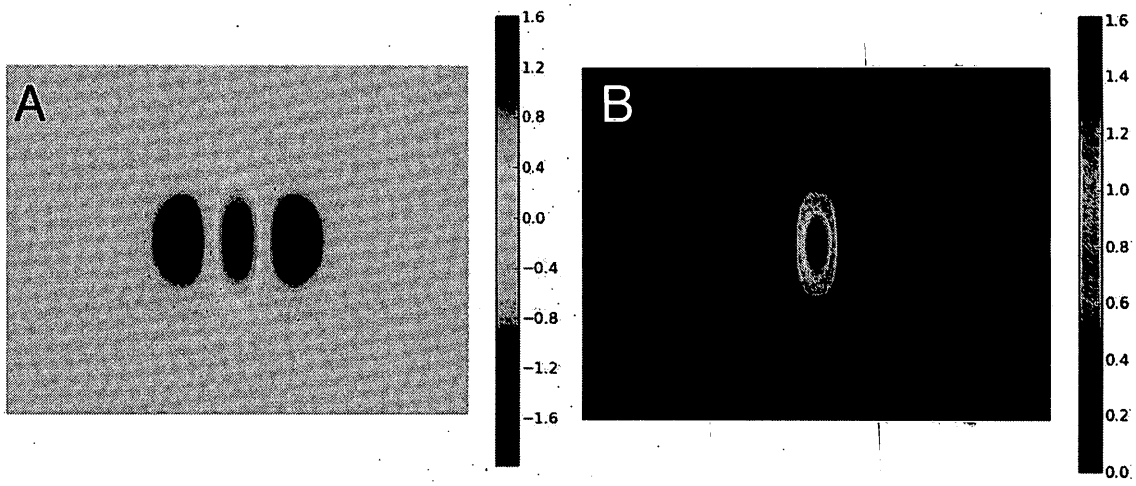

Figure 15. Multiple pRF estimates plotted in the visual space. Combining multiple Gaussians scaled with positively and negatively along a dimension of space creates edges and bands of activation and inactivation in the visual space (A). Thresholding the plot at 0 reveals a bar-like structure (B).

In the case illustrated in Figure 15, a series of negative Gaussians flanking either side of a series positive Gaussians will summate to a vertical bar.

A slight complication in the stimulus reconstruction procedure arises when one tries to relate the temporal dimension of the functional time series with that of the stimulus presentation. The stimulus reconstruction procedure here described attempts to use the voxel-wise pRF model estimates in tandem with the voxel-wise intensity fluctuations for the purpose of creating a facsimile of the visual stimulus. From frame to frame, the contribution of any given set of pRF estimates is determined via their scalar signal intensities. The hemodynamic response function describes the temporal dissociation between the stimulus and the response it evokes. Stimulus events occurring now have effective BOLD signals at some later point in time. The HRF is generally though to peak 4-6 s post-stimulus; however, the imprecision of this estimate might prove problematic for determining the signal intensity contribution of a given pRF estimate on the stimulus reconstruction.

The solution to the problem of identifying exactly *when* a stimulus occurring now manifests itself in the the unfolding BOLD signal of a given voxel is found via the explicit modeling of the HRF delay in the course of the pRF model estimation, aforementioned in section 3. Each pRF estimate includes both a three-parameter Gaussian as well as an HRF delay estimate. The delay describes the shift in the HRF peak with respect to onset. Since the delays derived using the model can be any real number but the ordinate dimension of

our functional datasets can only be integers, we have to interpolate the time series of each voxel. To do this, we used a cubic spline to up-sample the time series of the voxel 100 fold. In so doing, we are able to determine the signal intensity of a voxel at an onset delay of say 4.5 or 4.75 s rather than 4 or 5 s. The equation for scaling the Gaussian now becomes

$$g(x,y,\sigma,i) \quad = \quad exp - \left(\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma^2} \right) i_{t+d}$$

where i_{t+d} describes the signal intensity at a given time-point plus some derived HRF delay.

4.5 Stimulus reconstruction results

It is at this point that the stimulus reconstruction procedure can begin to demonstrate its effectiveness in rendering derivations of the stimulus presented to a participant based on the participants' voxel-wise pRF estimation maps and functional time series. Rather than using a single voxel or even a handful of voxels, the reconstruction procedure uses every voxel within a circumscribed region of interest (ROI) to render, scale, and summate the Gaussians. Doing so over large numbers of voxels across the entire functional time series yields stimulus reconstructions that approximate the stimulation pattern used during the visual presentation. Figure 16 shows a series of reconstructed frames from a

single participants' averaged dataset collected when using the sweeping bar aperture as the visual stimulus. Since the pRF estimates were derived from the same sweeping bar dataset that produced the stimulus reconstruction, this is less a cross-validation of the pRF model than a proof of concept for the stimulus reconstruction in general. Each column in Figure 16 represents a single time-point in the course of both the actual stimulus and the stimulus reconstruction.

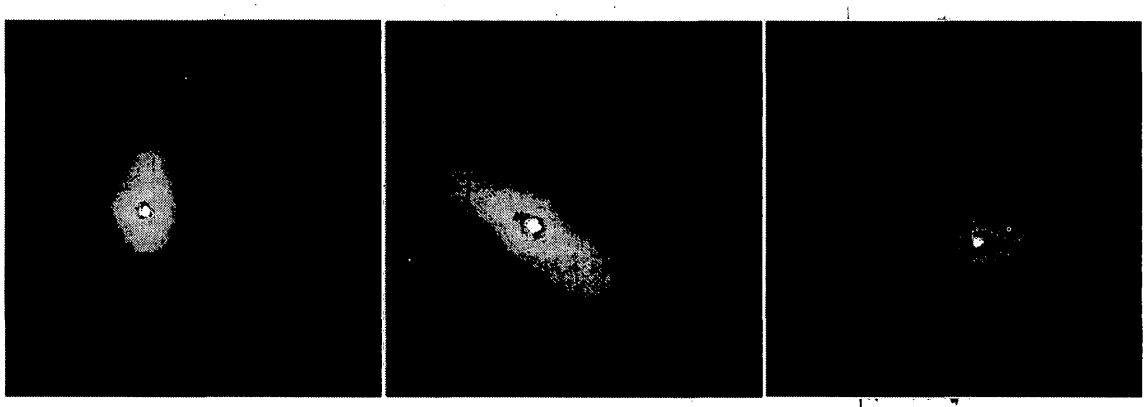

Figure 16. Binarized versions of the stimuli are shown on the top row at specific time intervals during the course of a stimulation run. The corresponding stimulus reconstructions are shown below on the bottom row. Each of the three columns correspond to a particular stimulus frame and its corresponding stimulus reconstruction frame. The stimulus reconstruction was generated using the mean run (computed from 5 runs) of one participant, with pRF estimates and signal intensities mined from voxels in V1, V2, and V3.

I've selected three frames for demonstration purposes, but it is important to note that these frames were chosen from a larger set of frames that comprises the entire duration of the visual presentation and functional dataset collection. The actual stimulus (the top row of Figure 16) and the stimulus reconstruction (the bottom row of Figure 16) are quite similar both in terms of the spatial configuration of pixel luminance within a single frame as well as the temporal continuity over the course of an entire run.

In order to cross-validate the stimulus reconstruction procedure, we used the pRF estimates derived from the sweeping bar stimulus and scaled these within the functional time series from other stimuli including an expanding ring and a rotating hemifield. The sweeping bars data, used to derive the pRF estimates, and the rotating hemifield data, used (in combination with the pRF estimates) to reconstruct the visual stimulus, for this particular participant were collected in different sessions—in fact, in different years. Each of these datasets were comprised of 8 32 s cycles. These data were originally designed and collected for deriving the retinotopy of early visual cortex using the phase-encoding approach described earlier. Nonetheless, these data serve as a good testbed for cross-validating our stimulus reconstructions.

Figure 17 shows the actual stimulus and the reconstructed stimulus for the rotating hemifield dataset from a single participant. A single mean cycle was computed from 7 of the 8 32 s cycles, having thrown away the first cycle due to the capitative nature of the BOLD signal. The voxels used in the reconstruction

were mined from visual areas V1, V2, and V3. Only those voxels with pRF estimates in which the covariance exceed 0.33 were accepted for candidacy in the reconstruction procedure. Again, three slices from the time axis were selected to demonstrate the apparent similarity between the actual stimulus and the reconstructed stimulus. The same procedure and selection process was used on the expanding ring data, shown in Figure 18.

As in the case of the sweeping bar and rotating hemifield, the similarity of the actual expanding ring stimulus and its reconstructed counterpart are obvious.

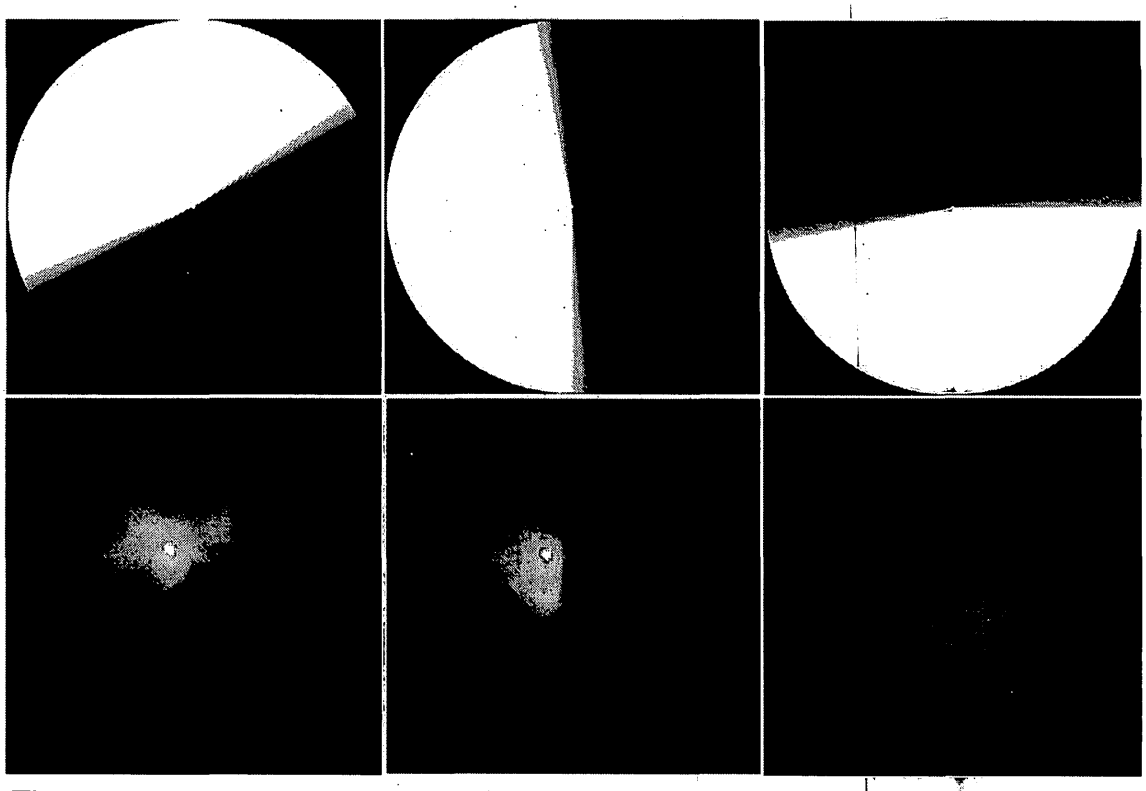

Figure 17. Frames selected from the binarized versions of the rotating hemifield stimulus and the corresponding stimulus reconstruction. The stimulus reconstruction was generated using the a mean cycle (32 s, 7 cycles) of a mean run (5 runs), with pRF estimates and signal intensities mined from voxels in V1, V2, and V3.

In all three cases, the spatiotemporal characteristics seem to be well-captured by the pRF model and stimulus reconstruction algorithm. That the sweeping bars are reconstructed faithfully shouldn't be very surprising; after all, the pRF estimates were derived from exactly these data. It is quite encouraging, though, that the reconstruction holds up so well with the phase-encoding stimuli, to which the algorithm was naive.

However, making judgments concerning the fidelity of the stimulus reconstructions based on the similarity in appearance of a handful of sampled

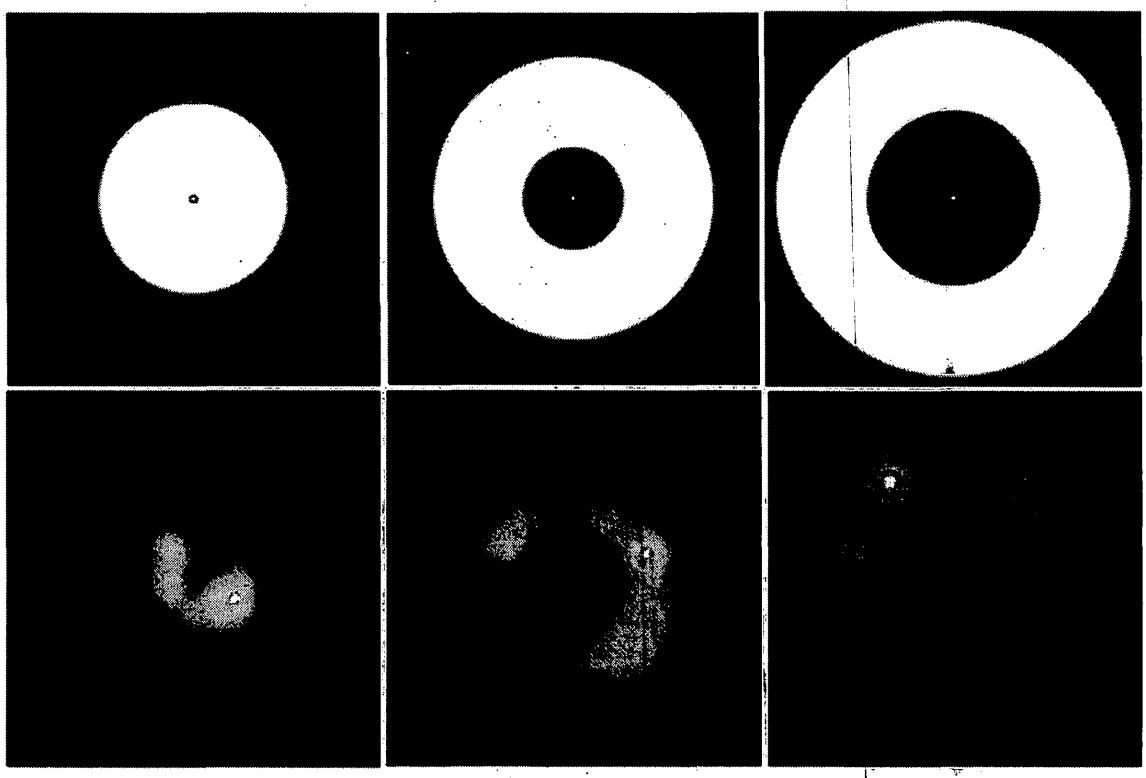

Figure 18. Frames selected from the binarized versions of the expanding ring stimulus and the corresponding stimulus reconstruction. The stimulus reconstruction was generated using the a mean cycle (32 s, 7 cycles) of a mean run (5 runs), with pRF estimates and signal intensities mined from voxels in V1, V2, and V3.

time slices isn't sufficient for rigorous evaluation. For quantitatively assessing the performance of the algorithm in reconstructing stimuli for novel functional datasets, it is necessary to use metrics that summarize the entirety of the spatiotemporal relationship between the stimuli and their reconstructions.

4.6 Stimulus reconstruction quantification

The stimulus reconstructions generated by the algorithm have several interesting properties that facilitate the attempt to decode the relationship between the measured neural activity and the pattern of visual stimulation. FIrst, every stimulus reconstruction exists in the same coordinate space. The spatial resolution of the stimulus reconstructions is always inherited from the visual display. This makes it trivial to combine data from multiple participants' once the stimulus reconstructions have been generated for a given functional dataset. Each participants' reconstruction generated from a given set of voxels will have a pixel-for-pixel (or degree-for-degree) correspondence with any other reconstruction, whether it be generated from a homologous set of voxels from another participant or from another set of voxels form the same participant.

Second, the spatiotemporal pattern of the reconstructions will vary depending on the set of voxels used as input to the algorithm. Various tasks engage different kinds of behaviors and will recruit different networks of brain regions. The variation in stimulus reconstruction as a function of voxel selection can be informative in terms of the kinds of information those voxels extract from

and represent about a stimulus or task. For instance, if a particular brain area responds primarily to the visual stimulation component of a task, the reconstruction should demonstrate a clear similarity in the luminance and contrasts of the stimulus. If, on the other hand, a brain region is primarily driven by the attentional component of a task, its reconstruction should be relatively unresponsive to the contrast and luminance components of the visual stimulus and instead should reflect the shifts in spatial attention around the visual field.

Third, since the reconstructions are generated via linear operations, the stimulus reconstructions that result from pooling the inputs (i.e., ROIs drawn in the brain imaging volume) or the outputs of the algorithm (i.e., stimulus reconstructions) are equivalent. For example, if one were interested in reconstructing the stimulus based on the pRF estimates and functional time series of V1 and V2, one could either add the regions of interest together and feed this into the reconstruction algorithm or add the individual reconstructions generated from V1 and V2 separately—although combining ROIs to create supersets of encompassing multiple areas is only sensible for a within-subjects analysis. Additionally, stimulus reconstructions can serve as operands for any other linear matrix operation, such as addition, subtraction, etc. Figure 19 shows the result of adding two different frames from the stimulus reconstruction of the expanding ring and rotating hemifield stimulus.

Fourth, since the stimulus reconstructions represent information about the spatiotemporal qualities of the stimulus in terms of pixel luminance fluctuations,

regions of interest can be defined in the screen coordinates from which a time-series can be extracted for the purpose of summarizing the pattern of activation. Figure 20 shows the mean time series derived from a single location within the display space of the stimulus reconstructions across multiple visual areas, including V1, V2, and V3. Figure 21 shows the result of the same kind of analysis, but this time using the rotating hemifield and expanding ring reconstructions instead of sweeping bar reconstruction. Here, the mean 32 s

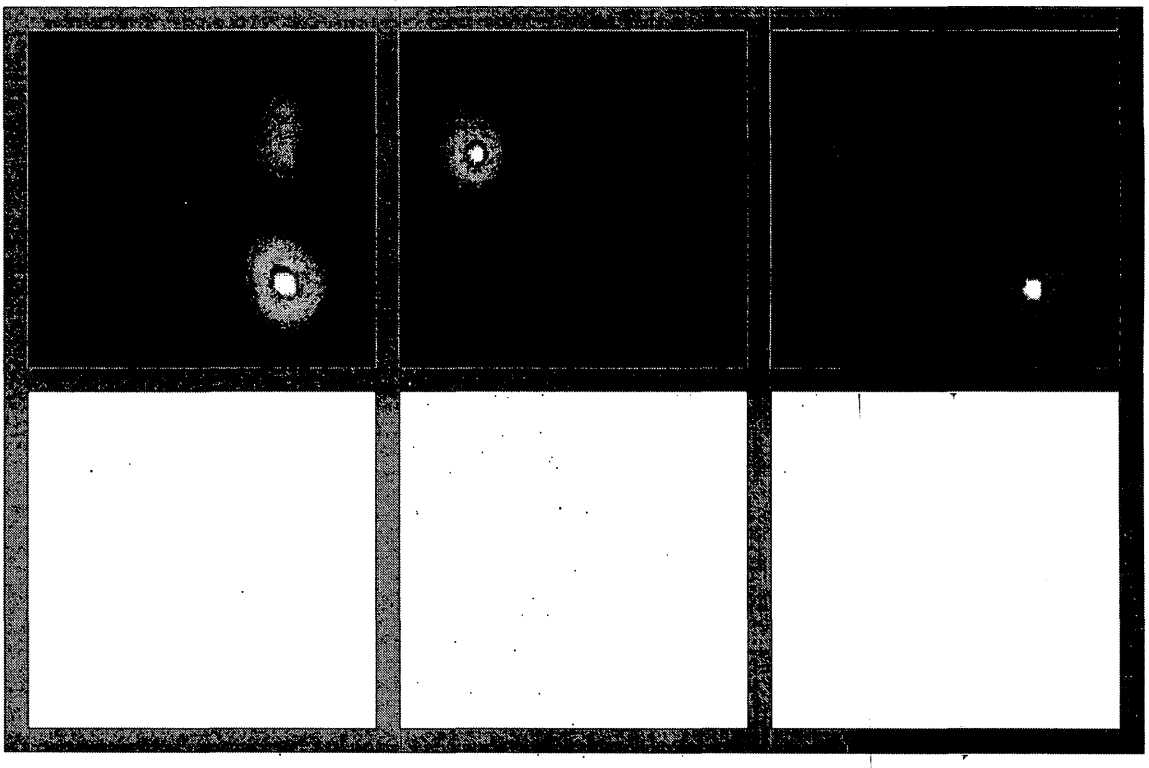

Figure 19. Stimulus reconstructions can act as operands in mathematical operations. Here, frames from the reconstruction of a bar stimulus extracted at different times are added together to show the response to two bars in the same frame simultaneously. Any linear matrix operation can be used to manipulate the content of the stimulus reconstructions, including addition, subtraction, and multiplication.

cycle is shown for the purposes of demonstration. Notice how the time-series derived from stimulus reconstruction are both sinusoidal.

4.7 Stimulus reconstruction discussion

The method of stimulus reconstruction as outlined here offers a new and interesting avenue for approaching and analyzing brain imaging data measuring visual processes. Ensembles of two-dimensional Gaussians can be linearly

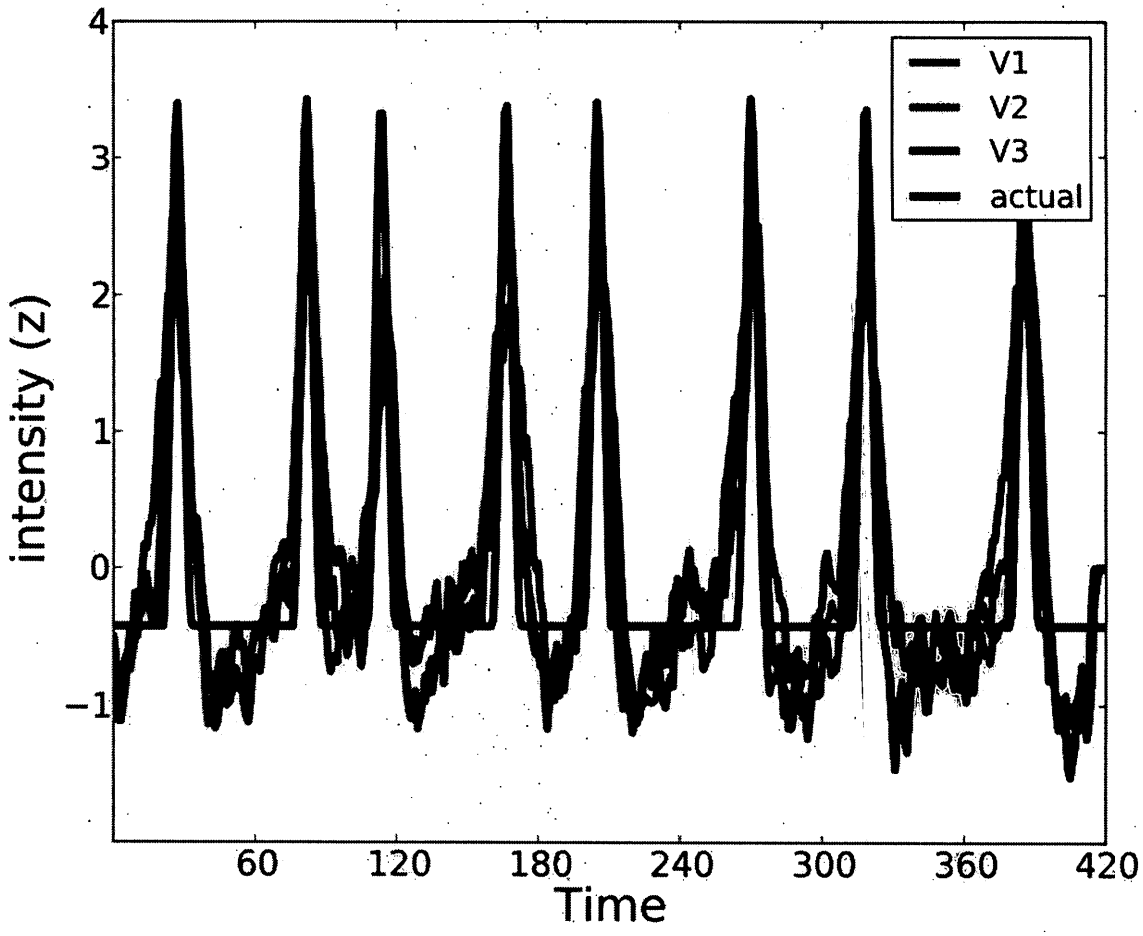

Figure 20. The mean luminance timeseries for stimulus reconstructions generated using V1, V2, and V3 ROI and the bars functional dataset. Each set of time-series across the visual areas was computed from pixels extracted from a particular location in the stimulus reconstruction. The time-series of the actual stimulus is shown in black.

combined (Figures 14 and 15) to produce new spatial configurations rich with information. Doing so across pRF estimates for variously defined clusters of voxels and scaled with the corresponding BOLD signal modulations produces stimulus reconstructions that mimic the spatiotemporal pattern of the visual stimulation (Figures 16, 17, and 18). Since the stimulus reconstructions are linearly related to the brain imaging data and since they exist in the coordinates of the visual space, time-series can be derived via screen coordinate masks that amass the pixel fluctuations of the stimulus reconstructions themselves (Figures 20 and 21).

In the context of the current thesis, the stimulus reconstructions can be used to derive the brain activity related to a particular location in the visual space and moment in time. This ability serves as a point of departure between the

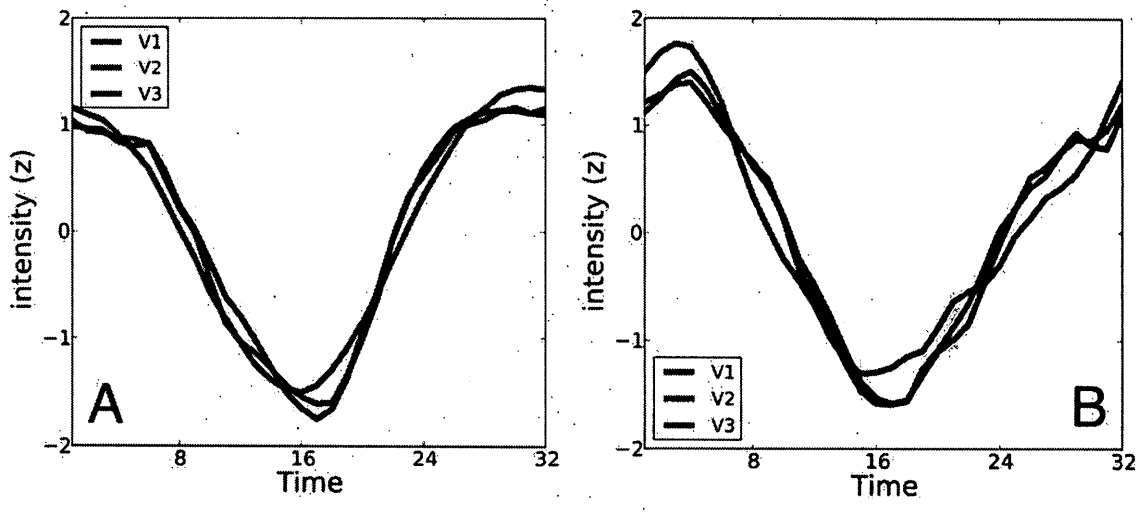

Figure 21. The mean luminance time-series for stimulus reconstructions generated using the rotating hemifield (panel A) and expanding ring (panel B) functional datasets. Here, the mean cycle of 32 s is computed for purposes of demonstration. The sinusoidal time-series of stimulus reconstruction is apparent.

current study and the majority of brain imaging experiments investigating tracking. Because we know the precise locations of targets and distractors over space and time as it was presented on the visual display, we can use this information to then refer to stimulus reconstructions at the very same locations in time and space and extract the corresponding brain activity. Thus, we are able to dissociate and compare signals related to attended versus unattended items and occluded versus nonoccluded item. Such an approach should serve as a model for further investigations into the spatiotemporal properties of retinotopic cortex in a variety of task contexts, including attention and occlusion.

5. Tracking

Having 1) computed the pRF maps for multiple participants and 2) developed a stimulus reconstruction method, we can now begin to explore the differential responses during a tracking task. We developed a much simpler version of the MOT task as outlined in previous experiments (Scholl & Pylyshyn, 1999; Flombaum et al., 2008). The reason for this was multifold and warrants some discussion.

Typically, the MOT task trial durations are under a minute. In addition, the initial positions of targets and distractors as well as the trajectories that unfold over a given trial are random by design so as to even sample the visual space. This poses a problem in terms of porting the MOT task to a brain imaging environment. Since the the signal-to-noise ratio of BOLD is relatively low, it is common to design experiments such that the stimuli and/or task are periodic over the course of scanning runs or sessions. Thus, the imaging data can be averaged across cycles, runs, or sessions so as to reduce the noise floor as much as possible. Using a randomly generated visual stimulus and task from trial to trial is prohibitive for combining datasets. Furthermore, most laboratories using the MOT task generate their stimuli such that the distribution of targets and distractors remains homogenous over the extent of the visual field from trial to trial. We made no such attempt as it was discovered *after* an alternate version of the task was adopted and because it would have been beyond the scope of the current project.

Attempts were made in the course of our experiment to use a 30 s tracking task. The starting positions and identities of the tracked items were reseeded to be the same at the start of every scanning run. Due to the deterministic nature of the motion equations governing how the items interacted with each other and the boundaries of the display through a trial, the trial would always unfold in the same way with the items always terminating their transit in the same positions. We then collected a multitude of 30 s brain imaging datasets and then computed the mean from these. The resulting stimulus reconstructions were so much noisy as chaotic. We abandoned this approach in favor of a scaled-down tracking task with on "target" and one "distractor". Participants maintained covert attention on the target via a monitoring task. Thus, a periodic design could be enforced and a mean cycle could be computed.

5.1 Tracking stimuli

The tracking stimuli consisted of two 2° white dots that transited around fixation on an orbit at 6° eccentricity and at a velocity of 2°/s. Each dot transited a full period around the orbit in 60 s, and each run consisted of 8 periods. The attention component was created via a rapid serial visual presentation (RSPV) was superimposed on each dot. Participants were instructed to count the occurrences of 'O' and 'X' among distractor letters and digits. The RSPV probes changed at a rate of 3 Hz. At the outset of run, the initial positions of the two dots

were set to the upper and lower vertical meridians and instructions were given to the participants to attend to the dot at the upper vertical meridian.

In addition to the dots and RSVP streams, an occluding bar was inserted into the display. The bar was oriented at 135° and spanned spanned fixation, extending to 10° eccentricity on either side of fixation. On half the runs, the northwest end of the bar occluded both the dot and the RSVP as it transited past. On the other half of the runs, the southeast end of the bar occluded the dot and the RSVP stream (Figure 22). Designing the tracking stimulus in this fashion allowed for the direct comparison of functional brain imaging signals corresponding to attended versus unattended and occluded versus non-occluded conditions at each of the two bar locations (135° polar angle, 6° eccentricity and 315° polar angle, 6° eccentricity).

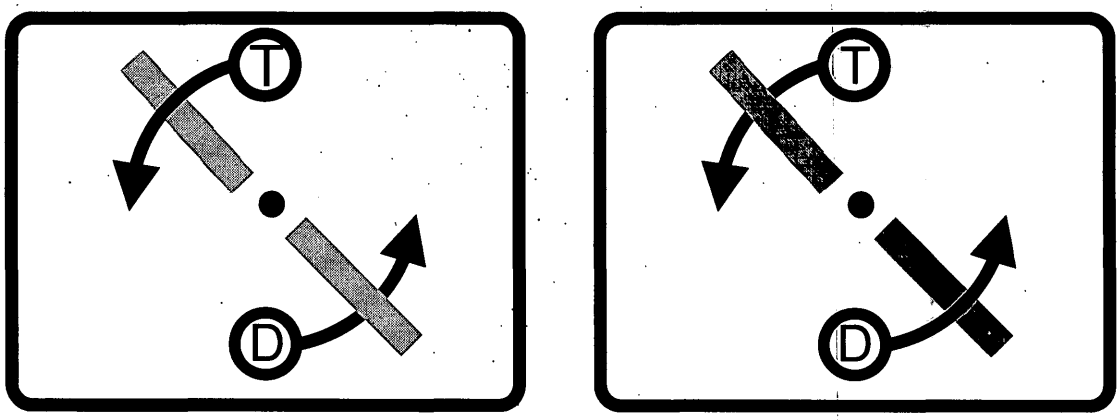

Figure 22. A schematic of the tracking experiment. At the outset of a run, two 2° wide dots with an RSVP superimposed on each appeared above and below fixation. Participants were instructed to attend the target and RSVP stream above fixation (T) and ignore the distractor and RSVP below (D). Each dot orbited fixation at a radius of 6°, witl an orbiting taking 60 s. Occluding bars were inserted into the task such that on half the runs the northwest bar occluded the attended and unattended RSVP streams (right panel) while on the other half of the runs the southwest bar occluded attended and unattended streams (left panel). Each orbit took 60 s and was repeated 8 times per run.

5.2 Tracking results

Stimulus reconstructions were generated for each participant and each visual area individually. Since the stimulus reconstructions exist in the common coordinate system of the visual space, combining datasets across participants is trivial. Figure 23 shows a single frame in time from the stimulus reconstruction generated using the orbiting stimulus. In this case, the frame taken shows a

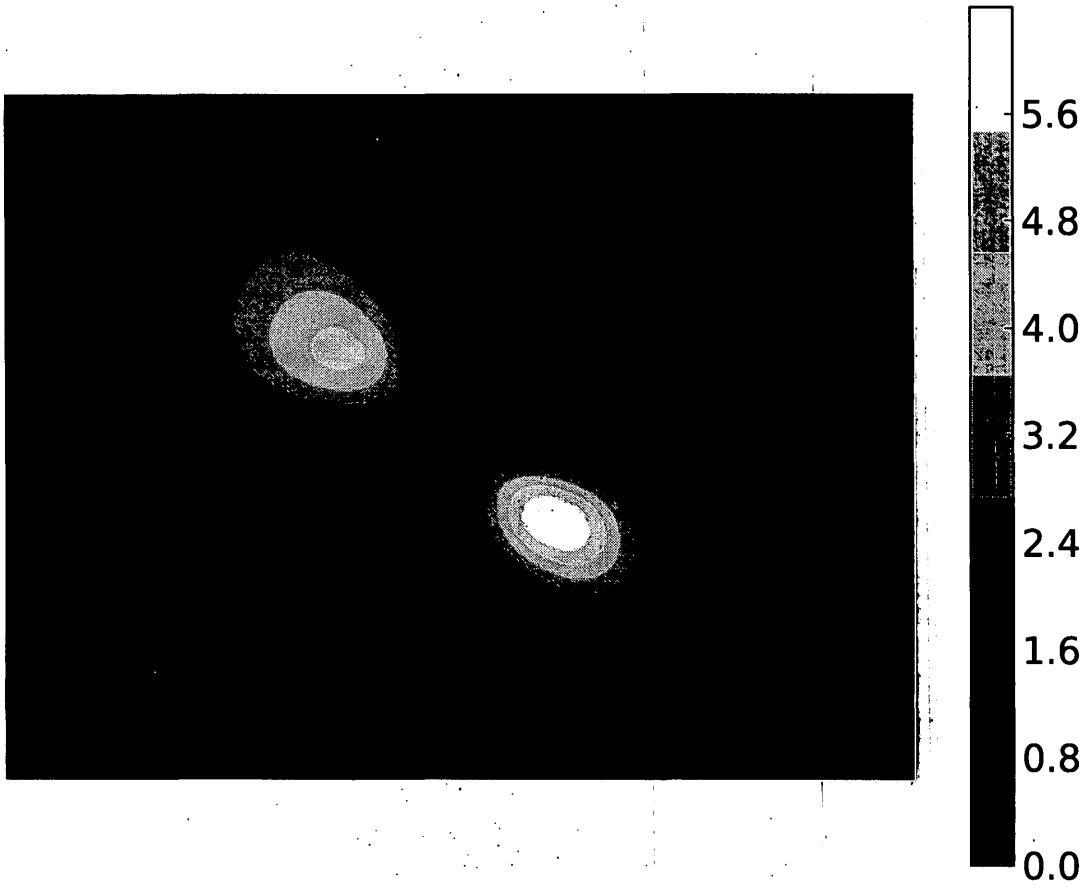

Figure 23. A frame from the stimulus reconstruction generated using the tracking stimulus. The peak in the upper left corresponds to the signal resulting from the nonoccluded distractor while the peak on the lower right corresponds to the signal resulting from the occluded distractor. Note how the representation of the target is higher and more narrowly tuned over space compared to the distractor. The scale is in z-units.

moment during the course of the mean cycle where the target is occluded by the bar in the southeast corner while the nonoccluded distractor is passing over the bar in the northwest corner. The stimulus reconstructions were variance normalized frame by frame so that the pixel intensities are expressed in z-units.

Typically, a region-of-interest brain imaging analysis computes a time-series based on the BOLD signal fluctuations over time within a circumscribed portion of brain tissue. Where it concerns stimulus reconstructions, though, the region-of-interest analysis can be conducted in terms of the coordinates of the display space. Because the locations of the the occluding and nonoccluding bars in the visual space were known, the mean time-series were computed by extracting and combining the time-course of groups of pixels at locations corresponding to the bars (two 2° diameter circular patches centered at 6° eccentricity and oriented at polar angles of 135° and 315°).

Figure 24 shows the time-series and associated metrics derived from the stimulus reconstructions generated using voxels from visual area V1 across four participants. In all panels, the colors of the lines and bars refer to the data extracted from different locations within the stimulus reconstructions as indicated in the legend. Solid lines and bars denote signals and metrics that were extracted from positions in the stimulus reconstructions related to the target item. Hollow lines and bars denote signals and metrics that were extracted from positions in the stimulus reconstructions related to the distractor item. The breaks in the lines at the midway point in the cycle were inserted for the sake of

clarity to demarcate when the target exited and the distractor entered a given position on the screen. From the perspective of pixels in the stimulus reconstruction corresponding to the orange box, the target passes through that location first followed by the distractor. Hence, at the beginning of the cycle, the orange line is solid as the target passes that location and then transitions to hollow as the distractor passes the same location later in the same cycle. The same color coding convention was used for the metrics shown in the bar plots. Note that comparison of time-series and bars colored orange or blue with those colored green or purple offers a glimpse at the differences in the BOLD signal related to occluded versus nonoccluded items. Again, the comparison between

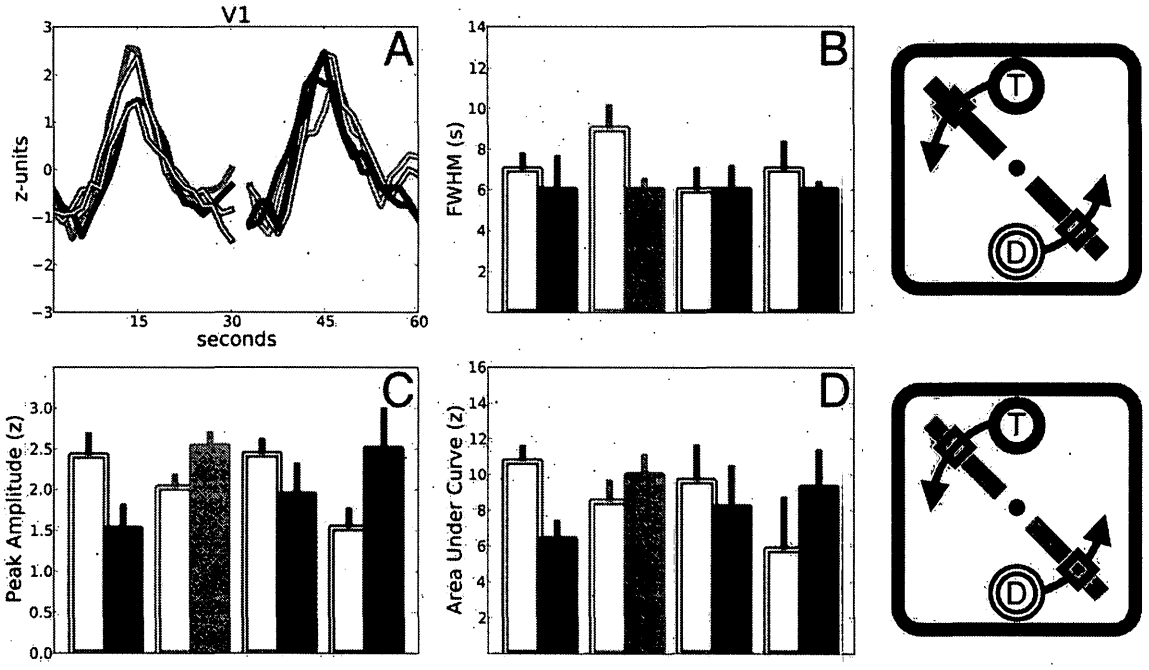

Figure 24. The results of the tracking experiment in visual area V1. All data was derived by computed the mean cycle across multiple runs and participants. Panel A shows time-series while panels B, C, and D show metrics summarizing the FWHM, peak amplitude, and area under the curve. The legend for all plots is displayed to the right. The signals are color-coded according to the point in the stimulus reconstruction from the data were extracted. Solid lines and bars denote signal and metrics related to the target (T). Hollow lines and bars denote signal and metrics related to the distractor (D).

solid and hollow lines and bars highlights the distinction in signal related to targets and distractors. The metrics in the bar plots were chosen for the purpose of conveying and scrutinizing the differences between the time-series across attention and occlusion conditions. We chose the peak amplitude, the area under the curve (AUC), and the full-width-half-maximum (FWHM) of each pair of peaks in the course of a full 60 s cycle. These metrics offer a summary of how fast the signal ramps up and down from the mean and by how much.

If we examine Figure 24 more closely, we can spot several interesting findings. Comparing the amplitudes associated with solid purple and orange to those amplitudes associated with the solid green and blue highlights the contrast

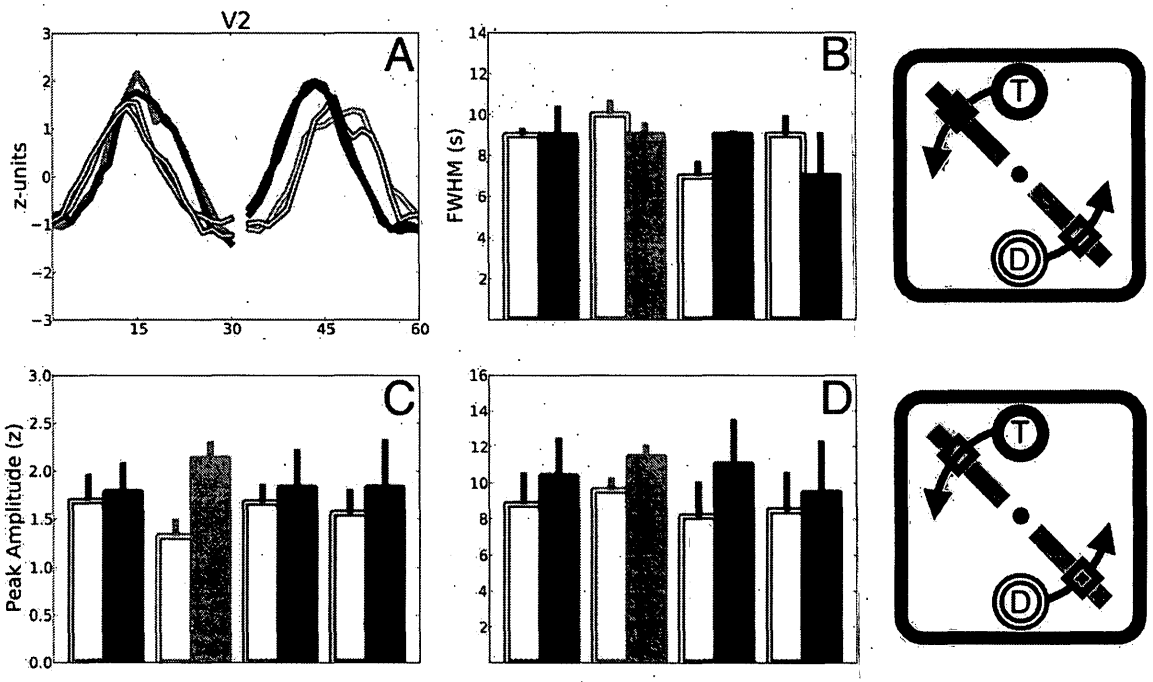

Figure 25. The results of the tracking experiment in visual area V2. All data was derived by computed the mean cycle across multiple runs and participants. Panel A shows time-series while panels B, C, and D show metrics summarizing the FWHM, peak amplitude, and area under the curve. The legend for all plots is displayed to the right. The signals are color-coded according to the point in the stimulus reconstruction from the data were extracted. Solid lines and bars denote signal and metrics related to the target (T). Hollow lines and bars denote signal and metrics related to the distractor (D).

in BOLD activity related to targets during occlusion versus nonocclusion. The results suggest that area V1 responds more strongly to targets during moments of occlusion than during moments when the target is nonoccluded. If we compare the hollow purple and orange with the hollow green and blue, we see that this pattern of results is reversed. Here, it appears as though area V1 responds more strongly to distractors during moments of nonocclusion compared to occlusion. If we compare these same conditions among the other metrics, the results indicate that targets during occlusion effect stronger but more brief BOLD responses compared to targets during nonocclusion. Distractors dominate the

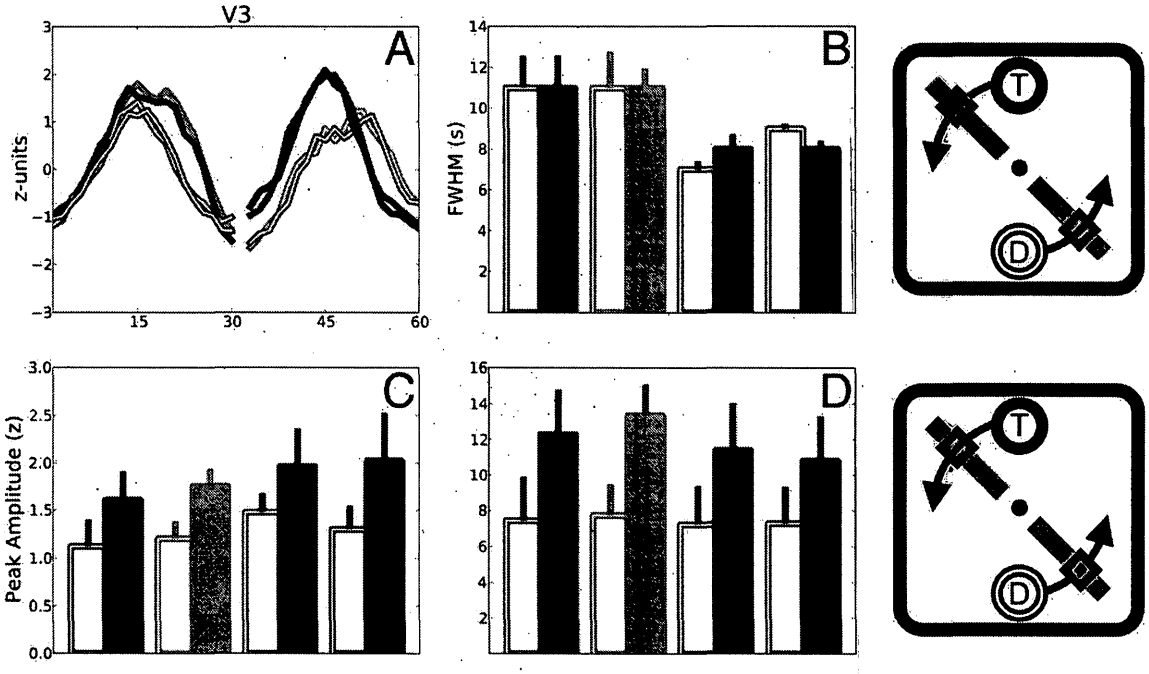

Figure 26. The results of the tracking experiment in visual area V3. All data was derived by computed the mean cycle across multiple runs and participants. Panel A shows time-series while panels B, C, and D show metrics summarizing the FWHM, peak amplitude, and area under the curve. The legend for all plots is displayed to the right. The signals are color-coded according to the point in the stimulus reconstruction from the data were extracted. Solid lines and bars denote signal and metrics related to the target (T). Hollow lines and bars denote signal and metrics related to the distractor (D).

BOLD signal at locations of nonocclusion (blue and green) and, as indicated by the AUC and FWHM metrics, seem to be more broadly tuned over space.

If we shift attention towards the same set of metrics extract from stimulus reconstructions generated using voxels from visual areas V2 (Figure 25) and V3 (Figure 26), the results become slightly more homogenous across occlusion and attention conditions. In both V2 and V3, the signals related to the target items—occluded or not—are higher than the signal related to the distractor items. An interesting break from this pattern is found in panel B of Figures 24 and 25. Here, the duration of the signal related occluded targets is less than the duration related to distractors at the same location.

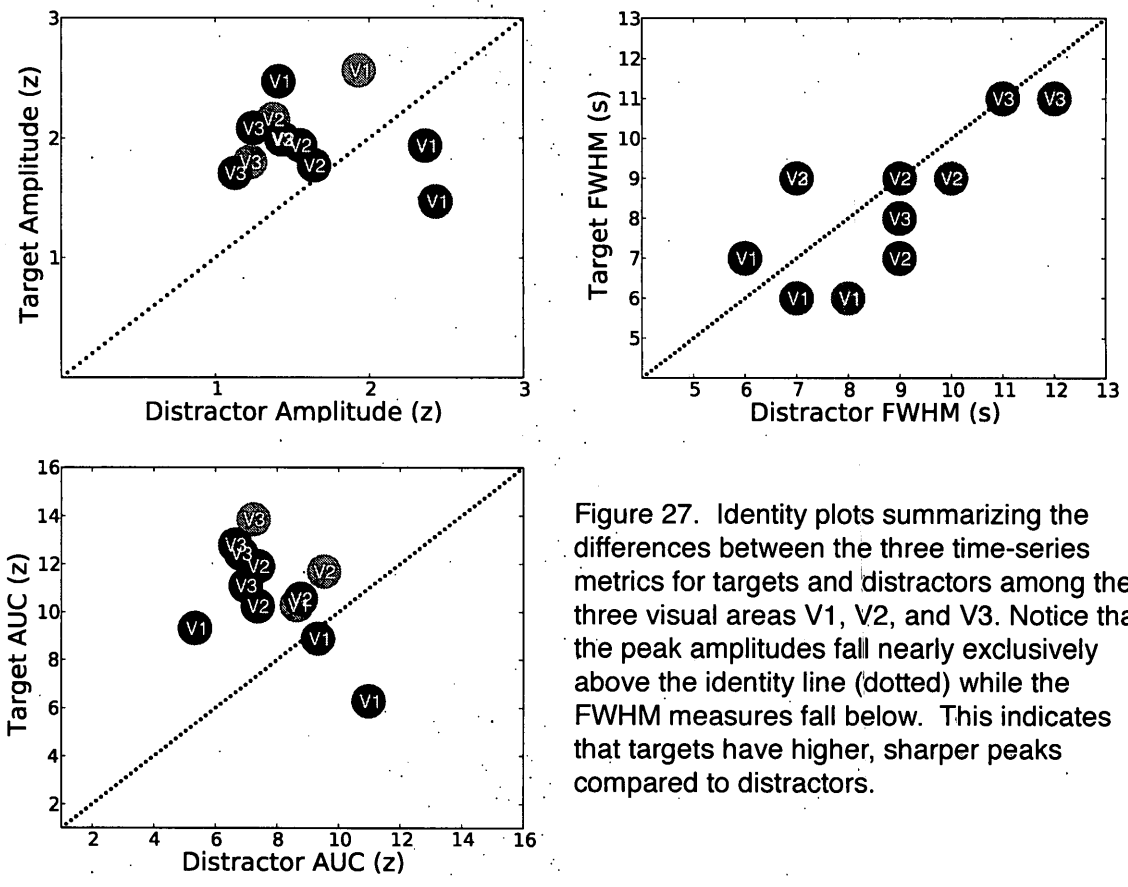

Figure 27. Identity plots summarizing the differences between the three time-series metrics for targets and distractors among the three visual areas V1, V2, and V3. Notice that the peak amplitudes fall nearly exclusively above the identity line (dotted) while the FWHM measures fall below. This indicates that targets have higher, sharper peaks compared to distractors.

The identity plots shown in Figure 27 summarize the effect of attention across all three visual areas and time-series metrics. In general, signals associated with targets are stronger and briefer compared to signal associated with distractors. This pattern is broken for nonocclusion in area V1 where the signals are stronger than their target counterparts.

5.3 Tracking discussion

The imaging findings reported here in response to a scaled-down version of a tracking task can be framed in terms of the psychophysical terminology revolving around the behavioral results of MOT occlusion studies (Pylyshyn, 2006; Flombaum et al., 2008). These experimenters noted that probes occurring on distractors are detected at a lower rate than probes occurring on targets. This result could manifest itself in terms of the fMRI data in one of two ways. If distractor inhibition is an active process whereby the regions of visual space surrounding distractors are actively suppressed so as do reduce distractor saliency, then probes occurring in these regions should be detected less often. If, on the other hand, regions of the visual space near distractors are simply ignored or otherwise under-sampled, then probes would presumably be missed more often when they appear near distractors. Each scenario—active suppressed distractor salience and passive non-enhanced distractor salience—would should have the same behavioral outcome (i.e., lower probe detection near distractors).

In terms of the BOLD signal, active distractor suppression should be met with an uptick in the local metabolic demand corresponding to regions of the retinomapic map at and around the distractor, albeit inhibitory, since increases in the BOLD signal related to local inhibition and excitation cannot be dissociated (Logothetis, 2008). The second possibility—that distractors are concomitant with regions of neglect in the visual space—should be met with little or no modulation in the BOLD signal. That we found found an increased BOLD activation in the regions surrounding *both* targets and distractors provides a neurophysiological basis for distractor inhibition.

In addition to finding a neurophysiological candidate for distractor inhibition, the results suggest that the BOLD signal reflects some amount of the high-beams effect reported previously in the behavioral literature (Flombaum et al., 2008). Here, probe detection is facilitated when they occur near targets *and* distractors during moments of occlusion compared to nonocclusion. The imaging results support this finding, but only for targets. The peak amplitude of occluded targets were higher than their nonoccluded counterparts. This was true across all three visual areas for both locations in the visual field where occlusion and nonocclusion events coincided in space. This effect did not hold for distractors, where nonoccluded distractors showed higher peak BOLD amplitudes across visual areas V1, V2, and V3. It is unclear why the effect should hold for targets but not for distractors, especially in light of Flombaum and others' finding that the relative occlusion advantage for probe detection was higher for distractors

compared to targets (although target probe detection was always higher in magnitude). The easiest explanation for the discontinuity between the behavioral and neurophysiological findings could be found in the differences between our task and their task. We used an overly simplified version of the MOT task and did not include any probe detection component. Participants were never cued to the distractor during the course of our runs, and so occlusion advantages at distractors may have been lost.

6. General Discussion

6.1 Technical challenges

The lion's share of work hours spent in the course of completing this thesis revolved around rewriting the pRF estimation routines and the entailing software support. The original authors' code-base for the pRF estimation (Dumoulin & Wandell, 2008) was written in MATLAB and designed primarily as a graphical user interface through which brain imagers could configure and run their analyses. Consequently, the pRF estimation code repository is quite byzantine. The author of this thesis wrote the pRF model estimation procedure anew using the open-source Python programming language (http://python.org), utilizing its various supporting scientific computing modules (e.g., http://www.scipy.org). In most cases, natively available tools were used for reading and writing files, performing time-series analysis, and model fitting. Where these tools did not exist, as in the case of the multi-scale adaptive brute-force estimation routine, the author created and programmed the algorithms. All software written by this author—including pRF estimation and stimulus reconstruction—will be incorporated into NIPY (http://nipy.sourceforge.net) and made available under a public use license.

6.2 Population receptive field estimation

The results of the pRF estimation experiment reconfirm retinotopic organization of early visual cortex using both the phase-encoding method (Engel et al., 1994; Sereno et al., 1995; DeYoe et al., 1996; Tootell et al., 1997) and the pRF model (Dumoulin & Wandell, 2008; Amano et al., 2009). Figures 6 and 7 demonstrate the orderly registry between visual field and the neuroanatomy. Based on the phase reversals of the polar angle maps, we were able to delineate the boundaries among cortical visual areas.

Although the scope of this study was limited to visual areas V1, V2, and V3, the pRF estimate maps from several subjects indicate a level of detail that would allow for the delineation of higher-order visual areas moving dorsally towards parietal cortex and ventrally and laterally towards temporal cortex. Several participants' pRF estimate maps were clean enough to suggest the emergence of visual areas along the dorsal pathway including areas V3a, V3b, and hV4 and several areas in IPS (Wandell et al., 2007). Additionally, a few participants' pRF maps showed retinotopically organized maps located laterally and ventrally to V3, including LO, VO, hV4, and MT+ (Wandell et al., 2007). That these areas weren't prominent among all participants is most likely attributable to lower signal-to-noise ratio of visual signals in these areas. The maps delineating the boundaries among these and even more areas are generally generated using multiple (i.e., more than 2) sessions of retinotopic mapping per participant. That

we were able to detect a subset of these areas with much fewer datasets is encouraging and warrants further investigation.

In addition to mapping more cortical regions, efforts are underway to implement the Difference-of-Gaussian (DoG) pRF model (Zuiderbaan et al., 2012). The version of the pRF model used in this thesis was the single Gaussian model, whereby the spatiotemporal reposne properties of each voxel are modeled as a 3-parameter Gaussian with single positive peak amplitude. This version of the model, however, is in capable of capturing the large negative BOLD response to the sweeping bars (see Figure 4, panels A and C for an illustration). Convolution of the effective stimulus with the double-gamma HRF—which includes a post-stimulus negative undershoot—allows for only subtle negative BOLD responses. In order to more accurately capture the variation of the BOLD about baseline, it is necessary to add an inhibitory surround to the pRF model. The resulting two-dimensional Gaussian is the so-called "mexican hat" function. Instead of a singular positive peak, the DoG model is function with an larger, central positive peak surrounded by an intermediate, negative peak. The DoG pRF model has been shown to explain more variance compared to the single Gaussian pRF model; in addition, the intercepts of relationship between eccentricity and receptive field size is lower across visual areas (Figure 11). For reasons unknown, the authors of the DoG pRF model use only center-on receptive field models and do not report any results using center-on *and* center-off receptive field models. An interesting extension of their work would include

using both center-on and center-off receptive field models for capturing visual signals in BOLD. Using the DoG model would also allow for more complex configurations of pRF estimates, potentially adding richness to the spatiotemporal information of the stimulus reconstructions.

6.3 Stimulus reconstruction

The stimulus reconstruction procedure outlined in this paper is novel in the field human brain imaging and shows a lot of promise for future mapping studies across sensory modalities. We demonstrated that stimulus reconstructions can be generated using the pRF estimates derived from the pattern of activation driven by the sweeping bar stimulus in combination with the BOLD signal fluctuations driven by any arbitrary pattern of visual stimulation. The algorithm works quite well for datasets used to train the pRF model (Figure 16) as well as datasets that were agnostic to the model estimation procedure (Figures 17, 18, and 23).

Furthermore, our stimulus reconstructions offer new avenues for approaching experimental design, data analysis, and the summarization of results. Since the stimulus reconstructions are linear combinations of voxel time-series data recast into the coordinate system of the display screen, combining datasets across visual areas and participants is trivial. The stimulus reconstructions themselves can serve as operands for any matrix operation including but limited to addition, subtraction, division, and multiplication (Figure

19). That the stimulus reconstructions exist in the coordinates of the display screen allows for region-of-interest analyses whereby the responses can be pooled over locations and extents of the display rather than of brain tissue. This affords an ability to inspect patterns of brain activation in terms of narrowly defined stimulus characteristics over time and space.

Since stimulus reconstructions are intended to be facsimiles of patterns of luminance fluctuations among pixels on a display screen, it is possible to use image similarity metrics to evaluate and tune the performance of the reconstruction algorithm. The depth of this project proved to be beyond the scope of the current thesis, but efforts are underway to implement a version of the algorithm whereby the contribution of voxels to the fidelity of the reconstruction can be determined. The approach taken is to iteratively increase and decrease the contribution of a given pRF estimate to the resulting stimulus reconstruction. If, for instance, amplifying the contribution of a given voxel to the stimulus reconstruction increases the dissimilarity between it and the actual stimulus, then the contribution of that voxel can be attenuated. However, because single frames from stimulus reconstructions are the result combining positively and negatively scaled Gaussians of various amplitudes, multiple voxel contribution configurations are possible. It is possible that settings among sets of voxels may be degenerate with each other, with no clear optimal solution. These are the sorts of issues that need to be worked out if any kind of optimization routine is implemented in the scheme of the stimulus reconstruction algorithm.

6.4 Tracking

The results of the tracking experiment seem to lend some neurophysiological support to the findings of MOT tasks using occlusion. We showed that the representations of targets across the visual field are higher and in amplitude and shorter in duration compared to distractors. Said another way, the responses to moving targets are more narrowly tuned over space. This effect is generally exaggerated during moments of occlusion. Distractors, on the other hand, are associated with weaker more slowly evolving signals. This would suggest that distractors are diffusely represented in the visual space and are broadly tuned over space. These findings indirectly suggest that the behavioral distractor inhibition results from an active suppression of rather than some kind of passive neglect of the neighboring visual space. A confound in our experiment is that we didn't include a control condition where purely visual signals are measured in response to an orbiting visual stimulus without the presence of a competitor. From this, we could compare whether the increase in BOLD signals associated targets and distractors could be assessed in terms of deviation from the stimulus-bound, tracking-free baseline. The stimulus reconstruction data structure would be ideal for conducting such a control experiment. Participants could passively fixate while untracked visual items transit around the visual field on trajectories identical to the tracking task. These data could be used to generate stimulus reconstructions that could be subtracted from those generated using the tracking

data. This would remove the signal attributed to the visual stimulus alone, and so we could make greater claims about the signals that emanate from the targets and distractors as attentional entities. As it is, the attention- and stimulus-drive signals are conflated.

Assuming for the moment that the signals related to distractors are indeed a form of neurophysiological distractor inhibition, the question now becomes: why should the visual system actively represent distractors at all? After all, the passive neglect explanation for distractor inhibition seems to be the simpler answer to the question of how to bias salience towards attended locations in the visual field. It may be the case that the goal is to both bias towards targets and away from distractors, but this explanation seems tautological. A more interesting explanation can be found in terms of recent findings in the neuroscientific literature that suggest a role of prominence for inhibitory signals in the early visual cortex of awake vertebrates. Haider and others (2013) found that cortical responses to visual stimuli in awake mice are the product of strong synaptic inhibition signals tuning the concomitant but weaker excitatory responses. The response to visual stimuli in awake mice was found to have a high ratio of inhibitory to excitatory signals and were narrowly tuned in space and time. Comparing the cortical output and the relative ratio of inhibition and excitation measured in awake mice in response to a visual stimulus, signals measured during anesthetize have a lower ratio of inhibitory to excitatory signals and are more broadly broadly tuned over time and space. That mixtures of

excitatory and inhibitory signals can increase the spatial and temporal tuning characteristics of neurons has been shown in a variety of sensory modalities (Isaacson & Scanziani, 2011). Both feedfoward and feedback inhibitory neural circuits are thought to help shape the response properties of neurons, effectively enhancing their sensitivity to a particular coordinate within some feature space.

Returning again to the tracking results, these findings shed some light on why the visual system bothers representing at all. While the BOLD signal is unable to dissociate local excitation from inhibition, it stands to reason that the explicit representation of targets and distractors would have the effect of spatially sharpening each of the items in the retinotopic space. Moreover, anytime there was a close encounter between targets and distractors in the visual field, neurophysiological distractor inhibition would presumably sharpen the boundaries of the target in both time and space. The task as it was conceived in our experiment does not afford us the ability to search for this sort of distractor-target spatial tuning interaction as they were separated by 180° of polar angle and 12° of eccentricity in the visual field. In light of both our findings and those of Haider and others (2013) warrant further investigation using a task where target and distractor items intersect in space.

That we found differences in BOLD activation in relation to attention and occlusion among our visual areas indicate that there is some kind of disparity in the types of information processing going on along the visual stream. Relating differences in the BOLD in relation to task demands with functional properties of

visual areas, though, is a complicated affair. The BOLD signal is considered to be associated with the inputs into a cortical region as well as the processing of that input by the local cortical circuitry (Logothetis, 2003). An increased response would suggest that the neural mechanisms have selected on a particular feature in the stream of visual input and are persisting and further differentiating that feature. If we turn again to the finding of increased BOLD signal for nonoccluded distractors in V1 but not V2 and V3, the results in tandem with the nature of BOLD would suggest that V1 contributes to the individuation of distractors in a particular attentional state. This individuation of nonoccluded distractors disappears upon inspecting the pattern of activation in areas V2 and V3. This train of thought could be continued for various aspects of the results in an effort to localize task demands in terms of the functional and anatomical hierarchy of the visual system. However, in light of the irregularity of the findings, more data should be collected and more experiments executed in order to determine the true neurophysiological order of operations in extracting and abstracting the various task demands related to tracking and occlusion.

References

Alvarez, G., & Franconeri, S. (2007). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. Journal of Vision, 7(13):14, 1-10.

Amano K, Wandell BA, Dumoulin SO. (2009) Visual field maps, population receptive field sizes, and visual coverage in the human MT+ complex. Journal of Neurophysiology. 102:2704-2718.

Beckmann, C.F. & Smith, S.M. (2002) Probabilistic independent component analysis in fMRI. Proceedings of the International Society of Magnetic Resonance in Medicine, 2002.

Bialek, W., Rieke, F., deRuyter van Steveninick, R.R., Warland, D. (1991) Reading a neural code. Science 252: 1854-1857.

Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J., 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. Journal of Neuroscience 16:4207–4221.

Carrasco M., Penpeci-Talgar C. & Cameron E.L. (2001) Characterizing visual performance fields: Effects of transient covert attention, spatial frequency, eccentricity, task and set size. Spatial Vision 15: 61-75.

Corbett, J.E., & Carrasco, M. (2011). Visual performance fields: Frames of reference. PLoS ONE 6(9):1-10.

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.

Culham, J. C., Cavanagh, P., & Kanwisher, N. G. (2001) Attention response functions: Characterizing brain areas using fMRI activation during parametric variations of attentional load. Neuron 32: 737-745.

Dale, A.M., Fischl, B., Sereno, M.I. (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. NeuroImage 9:179-194.

DeYoe, E.A., Carman, G.J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., Miller, D., Neitz, J. (1996) Mapping striate and extrastriate visual areas in human cerebral cortex. Proceedings of the National Academy of Science 93:2382–2386.

Dumoulin SO, Wandell BA. (2008) Population receptive field estimates in human visual cortex. NeuroImage 39(2):647-660

Engel, S.A., Rumelhart, D.E., Wandell, B.A., Lee, A.T., Glover, G.H., Chichilnisky, E.J., Shadlen, M.N. (1994) fMRI of human visual cortex. Nature 369, 525.

Engel S.A., Glover, G.H., Wandell, B.A. (1997) Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cerebral Cortex 7:181–192.

Fischl, B., Sereno, M.I., Dale, A.M. (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. NeuroImage 9: 195-207.

Fletcher, R., Powell, M.J.D. (1963) A rapidly convergent descent method for minimization. Compututation Journal 6: 163–168.

Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R. (1998) Event-related fMRI: characterizing differential responses. NeuroImage 7:30–40.

Glover, G.H. (1999) Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9: 416–429.

Haider, B., Hausser, M., Carandini, M. (2013) Inhibition dominates sensory responses in the awake cortex. Nature, 493:97-102.

Hartline, H.K. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. American Journal of Physiology 121:400-415.

Harvey BM, Dumoulin SO. (2011) The relationship between cortical magnification factor and population receptive field size in human visual cortex: constancies in cortical architecture. Journal of Neuroscience 31: 13604-13612.l

Holmes G. (1918) Disturbances of vision by cerebral lesions. British Journal of Ophthalmology 2:353-384.

Howe, P., Horowitz, T., Morocz, I., Wolfe, J., & Livingstone, M. (2009). Using fMRI to distinguish components of the multiple object tracking task. Journal of Vision 9(4):10.

Hubel D.H. and Wiesel T.N. (1962) Receptive fields, binocular interaction, and functional architecture of cat striate cortex. Journal Physiology (London) 160:106-154.

Hubel D.H. and Wiesel T.N. (1968) Receptive fields and functional architecture of monkey striate cortex. Journal Physiology (London) 195:215-243.

Isaacson, J.S., Scanziani, M. (2011) How inhibition shapes cortical articity. Neuron, 72:231-242.

Flombaum, J. I., Scholl, B. J., & Pylyshyn, Z. W. (2008). Attentional resources in tracking through occlusion: The high-beams effect. Cognition 107: 904-931.

Jovicich, J., Peters, R., Koch, C., Braun, J., Chang, L., & Ernst, T. (2001) Brain areas specific for attentional load in a motion-tracking task. Journal of Cognitive Neuroscience 13: 1048-1058.

Kanwisher, N. & Wojciulik, E. (2000) Visual attention: insights from brain imaging. Nature Reviews Neuroscience 1:91-100.

Kamitani, Y, Tong, F. Decoding the visual and subjective contents of the human brain.Nature Neuroscience, 8:679-685.

Kastner S., DeWeerd P., Desimone R., Ungerleider L.G. (1998) Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. Science 282: 108-111.

Kastner S., DeSimone K., Konen C., Szczepanski S., Weiner K., Schneider K. (2007) Topographic Maps in Human Frontal Cortex Revealed in Memory-Guided Saccade and Spatial Working Memory Tasks. Journal of Neurophysiology 97(5): 3494-3507.

Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L. (2008) Identifying natural images from human brain activity. Nature 452: 352:356.

Power J.D., Barnes K.A., Snyder A.Z., Schlaggar B.L., Petersen S.E. (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage 59:2142–2154.

Pylyshyn, Z. W., & Storm, R. W. (1988) Tracking multiple independent targets: Evidence for a parallel tracking mechanism. Spatial Vision 3:179-197.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial index model. Cognition 32:65-97.

Pylyshyn, Z.W. (2006) Some puzzling findings in multiple object tracking (MOT): II Inhibition of moving nontargets. Visual Cognition, 14:175-198.

Ren, D., Chen, W., Liu, C., & Fu, X. (2009). Identity processing in multiple-face tracking. Journal of Vision 9(5):18.

Saad, Z.S., Glen, D.R, G Chen, G., Beauchamp, M.S., Desai, R., and Cox, R.W. (2009) A new method for improving functional-to-structural alignment using local Pearson correlation. NeuroImage, 44:839-848.

Scholl, B. J., Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. Cognitive Psychology, 38:259–290.

Schneider K.A. 2011. Subcortical mechanisms of feature-based attention. Journal of Neuroscience 31: 8643–8653.

Sereno, M.I., A.M. Dale, J.B. Reppas, K.K. Kwong, J.W. Belliveau, T.J. Brady, B.R. Rosen, and R.B.H. Tootell (1995) Borders of multiple visual areas in

human revealed by functional magnetic resonance imaging. Science 268:889-893.

Stanley, G.B., Li, F.F., Yang, D. (1999) Reconstruction of natural scenes from ensemble responses in the geniculate nucleus. Journal of Neuroscience 19 (18): 8036-8042.

Treisman, AM, Gelade, G (1980) A feature-integration theory of human attention. Cognitive Psychology 12:97-136.

Thomas CG, Harshman RA, Menon RS. (2002) Noise Reduction in BOLD-Based fMRI Using Component Analysis. NeuroImage. 17:1521–1537.

Tootell, R.B.H., J.D. Mendola, N.K. Hadjikhani, P.J. Ledden, A.K. Liu, J.B. Reppas, M.I. Sereno, and A.M. Dale (1997) Functional analysis of V3A and related areas in human visual cortex. Journal of Neuroscience 17:7060-7078.

vanMarle, K., & Scholl, B. J. (2003). Attentive tracking of objects vs. substances. Psychological Science 14:498-504.

Wandell B.A., Dumoulin S.O., Brewer A.A. (2007) Visual field maps in human cortex. Neuron. 56(2): 366-383.

Wojciulik, E., & Kanwisher, N. (1999) The generality of parietal involvement in visual attention. Neuron 23: 747-764.

Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G.H., Morales, F., Evans, A.C. (2002) A general statistical analysis for fMRI data. NeuroImage 15:1–15.

Zuiderbaan, W., Harvey, B., Dumoulin, S.O. (2012) Modeling center-surround configurations in population receptive fields using fMRI. Journal of Vision 12 (3):10, 1-15.

Zwillinger, D. and Kokoska, S. (2000). CRC Standard Probability and Statistics Tables and Formulae. Chapman & Hall: New York.