From Photons to Photos: Mapping Functional and Organizational Properties of Human Visual Cortex with fMRI

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the Graduate School of The Ohio State University

By
Daniel E. Berman B.S.
Graduate Program in Psychology
The Ohio State University
2015

Master's Examination Committee:
Julie Golomb – Advisor
Dirk Bernhardt-Walther
Zhong-Lin Lu
Abstract

Immense strides in furthering our understanding of a variety of underlying mechanisms involved in visual perception are continuously being made, however numerous questions in the field remain unanswered. Detailed here are two studies aimed at elucidating the functional processes in which the human visual system integrates visual stimuli into cohesive representations of the world around us. First, using simulations and fMRI experiments we examined how retinotopic mapping paradigms can be optimized to most efficaciously and accurately measure the underlying functional and organizational properties of early to mid-level visual cortex in both healthy and clinical populations. We focused on quantitatively comparing two common retinotopic mapping designs in terms of measurable and achievable signal-to-noise ratios under various conditions such as exogenous noise and scanning time. While multifocal retinotopy can yield a higher signal-to-noise ratio in primary visual cortex, we find that the phase-encoded technique provides qualitatively better retinotopic maps despite having a less robust point-to-point correspondence of visual field location to visual cortex. Other implications of employing these methods are also discussed.

Next, fMRI experiments aimed at uncovering the spatial frequency tuning profiles of higher-level, category selective areas of cortex such as the parahippocampal place area (PPA), retrosplenial cortex (RSC), occipital place area (OPA), lateral occipital cortex
(LOC) and the fusiform face area (FFA) were performed. These experiments utilized both frequency filtered scenes belonging to four different categories along with unstructured stimuli of various spatial frequencies. Results from univariate, multivariate and searchlight analyses are subsequently provided and elaborated on, with the general finding that the category selective areas mentioned above are sensitive to high spatial frequencies, especially when semantic information is present.
Acknowledgements

Special thanks to project collaborators Zhong-Lin Lu, Xiangrui Li and my advisors Dirk Bernhardt-Walther and Julie Golomb for their many contributions and insights pertaining to this work.

This work was supported by The Ohio State University, Columbus, startup funds of Dirk Bernhardt-Walther.

All fMRI scanning was performed in the Center for Cognitive and Behavioral Brain Imaging (CCBBI) facilities.
Vita

May 2006........................ Oakland Mills High School

May 2010........................ B.S. Psychology, with Honors, University of Maryland

Fields of Study

Major Field: Psychology
Table of Contents

Abstract.............................................................................................................................................. ii
Acknowledgments ............................................................................................................................... iv
Vita....................................................................................................................................................... v
List of Tables ....................................................................................................................................... vii
List of Figures ..................................................................................................................................... viii
General Introduction .......................................................................................................................... 1
Chapter 1: Optimizing fMRI retinotopic mapping methodologies ..................................................... 5
Chapter 2: Spatial frequency tuning profiles of category selective visual areas.............................. 53
References .............................................................................................................................................. 92
Appendix A: Chapter 1 Supplementary information ........................................................................ 102
Appendix B: Chapter 2 Supplementary information ........................................................................ 107
List of Tables

Table 3.1) Simulated SNRs for phase-encoded and multifocal mapping paradigms........27
Table B.1) Decoding Accuracies - Scene Category (4 Scenes).................................109
Table B.2) Decoding of Spatial Frequencies - (Binary) Scenes and Unstructured
Stimuli......................................................................................................................110
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Phase-encoded stimuli</td>
<td>9</td>
</tr>
<tr>
<td>2.1</td>
<td>Autocorrelation of an m-sequence</td>
<td>16</td>
</tr>
<tr>
<td>2.2</td>
<td>Four different multifocal stimuli</td>
<td>18</td>
</tr>
<tr>
<td>2.3</td>
<td>Additional examples of multifocal stimuli</td>
<td>19</td>
</tr>
<tr>
<td>2.4</td>
<td>Recreations of the stimuli used in Pihlja et. al. (2008)</td>
<td>20</td>
</tr>
<tr>
<td>3.1</td>
<td>Design matrix of the phase-encoded stimuli</td>
<td>26</td>
</tr>
<tr>
<td>3.2</td>
<td>Spatial relations proximity matrix</td>
<td>28</td>
</tr>
<tr>
<td>3.3</td>
<td>Multifocal and phase-encoded SNRs by noise level (all voxels)</td>
<td>29</td>
</tr>
<tr>
<td>3.4</td>
<td>Interaction between noise level and experiment length</td>
<td>30</td>
</tr>
<tr>
<td>3.5</td>
<td>Simulated SNR according to experiment length</td>
<td>31</td>
</tr>
<tr>
<td>4.1</td>
<td>Example multifocal stimulus used in fMRI experiment</td>
<td>37</td>
</tr>
<tr>
<td>4.2</td>
<td>Test-retest reliability of multifocal retinotopic mapping</td>
<td>40</td>
</tr>
<tr>
<td>4.3</td>
<td>Multifocal and phase-encoded flat maps</td>
<td>42</td>
</tr>
<tr>
<td>4.4</td>
<td>Average SNR across visual areas – multifocal mapping</td>
<td>43</td>
</tr>
<tr>
<td>4.5</td>
<td>Receptive fields of 100 randomly selected voxels in V1</td>
<td>44</td>
</tr>
<tr>
<td>4.6</td>
<td>Receptive fields of approximately 1000 voxels in V1</td>
<td>45</td>
</tr>
<tr>
<td>4.7</td>
<td>Local cortical representations of visual field measured using the multifocal stimulus</td>
<td>47</td>
</tr>
</tbody>
</table>
Figure 5.1 Examples of the unstructured stimuli .................................................................62
Figure 5.2 Examples of a jointly contrast normalized beach scene .........................63
Figure 5.3 Sample experiment sequence ...........................................................................64
Figure 6.1 Percent signal change in each experimental condition .............................71
Figure 6.2 Univariate results of PPA activation levels (L/R hemisphere; A/P regions) 73
Figure 6.3 Decoding accuracy - scene category .................................................................76
Figure 6.4 Decoding scene category - PPA and A/P subregions of PPA ...............78
Figure 6.5 Decoding scene category within PPA by hemisphere ...............................80
Figure 6.6 Decoding spatial frequencies - PPA and A/P subregions of PPA ..........82
Figure 6.7 Searchlight analysis - PPA ..............................................................................84
Figure 6.8 Searchlight analysis - OPA .............................................................................84
Figure A.1 Graphical depiction of canonical HRF function - chapter 1 ...............103
Figure B.1 Graphical depiction of canonical HRF function - chapter 2 ...............108
General Introduction

Examining the human visual system and the functional roles of specific cortical areas that give rise to representations of complex visual stimuli has been an enduring challenge for researchers across numerous disciplines. This interdisciplinary endeavor has yielded a bounty of information on the underlying neural processes that take place within the visual system and there is no doubt that we do understand an immense amount about the nature in which elementary particles (i.e. photons) are absorbed by retinal cells in the eye and computationally transformed into colors, objects, surfaces and all the necessary information that allow us to make sense of the world around us. Examining the hierarchical organization of the human visual system and the functional properties along this visual stream are fundamental in understanding this topic.

The starting point of the visual processing hierarchy is of course the eye, which contains retinal ganglion cells that absorb light and initiate the underlying mechanisms of sight. This information is transferred along the optic nerves of each eye, which meet and cross at the optic chiasm where visual field information is split and sent along optic tracts that terminate at the lateral geniculate nucleus (LGN). From the LGN, information is then relayed to primary visual cortex, known as V1. V1 cells have very specialized tuning to extremely low level visual properties like orientation, contrast and spatial frequency. From here a series of computations is performed that relay information to subsequent areas along the visual stream, such as V2, V3, and V4 where these low level
properties are further processed and eventually sent to higher level visual areas that can finally integrate this information into coherent representations of the world (Felleman & Van Essen, 1991; Van Essen & Gallant, 1994). High-level visual areas often have strong selectivity for certain stimuli and these implications will be discussed, but first it is important to discuss some more specific organizational properties of the visual system.

Each of the areas mentioned in the preceding whirlwind overview of the visual hierarchy and steps along the visual stream is organized retinotopically. This retinotopic organization refers to the fact that visual input hitting the retina is sent along the retinogeniculo-striate pathway so that adjacent areas of visual field are mapped onto adjacent locations of visual cortex (Guido, 2008; Wandell, 1995). This alone is an important consideration when examining the visual system, but the most essential implication of the retinotopic mapping of visual input is that even at the earliest processing stages there are differences in functional properties that can be recognized amongst these maps. For example, within the LGN there are two main pathways that lead to primary visual cortex, the magnocellular and parvocellular pathways. It has been observed that high spatial frequency information is processed more by the parvocellular pathway and that the magnocellular pathway processes low spatial frequency information (Derrington & Lennie, 1984) and that areas of visual cortex corresponding to lower eccentricities are tuned towards higher spatial frequencies and low spatial frequency tuning increases in cortical areas representing more peripheral areas of visual field (De Valois, Albrecht, & Thorell, 1982; Henriksson et. al., 2008). This is just one of the numerous examples of the relationship between organization and function underlying visual system processes, the
point being that in order to gain a full understanding of how complex representations of the visual world arise, it is also necessary to examine the retinotopic organization of visual cortex. This is the focus of the first study described below.

Building on these inquiries regarding the organizational nature of visual cortex, it is essential to examine the functional processes that allow for low-level information to be integrated. As described previously, early visual areas are processing low-level properties of images such as color, orientation, contrast, and spatial frequency. However, this is clearly not how we view the world. We visually perceive things not as an assortment of line orientations and colors; but as objects, places, and other forms that are sensible and have meaning. Recently, through the advent of fMRI, numerous high-level visual areas have been shown to play critical roles for integrating low-level sensory information into actual representations. For example, two areas located anterior to area V4 in the temporal lobe known as the parahippocampal place area (PPA) and fusiform face area (FFA) are selective to places (Epstein & Kanwisher, 1998) and faces (Kanwisher, McDermott & Chun, 1997) respectively. Other areas such as retrosplenial cortex (RSC) and occipital place area (OPA) have also shown selectivity to buildings and places (O’Craven & Kanwisher, 2000; Dilks et. al., 2013) and lateral occipital cortex (LOC) has been shown to be selective to objects (Malach et. al., 1995).

The relationship of low-level visual properties with these high-level areas is still not well understood. One low-level property that several studies have either directly or indirectly looked at with respect to these areas is spatial frequency. For example, Rajimehr et. al. (2011) reported that scene selective area PPA responds more to higher
spatial frequencies. Other studies show PPA to be sensitive to global properties of images that are determined by low spatial frequencies such as openness (Park et. al., 2011) as well as physical size and clutter (Park, Konkle & Oliva, 2014). These opposing findings, along with the absence of knowledge pertaining to the spatial frequency sensitivities of the other high-level visual areas mentioned above, are the motivation of the second study described below. The focus of the second study is elucidating the spatial frequency tuning characteristics of these areas in hope of furthering our understanding of the nature in which low-level properties of images are integrated to high-level representations.
Chapter 1: Optimizing fMRI retinotopic mapping methodologies

Introduction

The retinotopic organization of visual cortex has been a topic of great interest for researchers and clinicians ever since neurologists during World War I established the correlation between isolated brain lesions and the locations of visual field loss observed in soldiers (Holmes, 1918; Fishman, 1997). Recently, the use of fMRI has exponentially increased our understanding of this organization throughout much of the visual system using a handful of established retinotopic mapping paradigms. The most frequently employed of these procedures include the phase-encoded (or traveling wave) method (Sereno et. al., 1995; Engel et. al., 1994; Sereno, McDonald & Allman, 1994), the population receptive field estimation method (Dumoulin & Wandell, 2008) and the multifocal method (Vanni, Henriksson & James, 2005). Being that fMRI research is a relatively new and burgeoning advancement in vision research, there is an ongoing desire to improve upon these methodologies in terms of efficiency, reliability, and spatial resolution. The first part of this manuscript will describe these techniques in detail and discuss their relative strengths and weaknesses. Additionally, the results of fMRI experiments and simulations designed to directly compare the efficacies of each paradigm will be detailed. Finally, recommendations regarding future research on
clinical populations with various visual impairments and the optimal retinotopic mapping approaches to utilize in these experiments will be provided.

**Purpose**

In order to provide the appropriate context for the explication of retinotopic mapping methodologies contained in this section, it is necessary to discuss the motivation for performing these analyses. Primarily, retinotopy provides the means for researchers to functionally define the boundaries of visual areas. Defining the boundaries of early visual areas is relatively straightforward because adjacent retinal neurons project to adjacent locations of V1 by way of the retino-geniculo-striate pathway, preserving the relative spatial correspondence of visual field and cortical organization (Guido, 2008; Wandell, 1995). In visual areas V2 and V3 visuospatial representations in cortex are not as consistent as in V1 but the maps are still quantifiable (Engel, Glover & Wandell, 1997; Dougherty et. al., 2003). Beyond these low-level visual areas retinotopic mapping is considerably more difficult for a variety of reasons, namely increased receptive field sizes and interfering neural connections (Smith et. al., 2001; Tsotsos, 2002), but researchers have still been able to establish numerous maps further down the visual stream (see Barton & Brewer, (2012) for a review of established retinotopic areas). These functionally based approaches for retinotopic mapping are an especially important achievement for vision based fMRI research because until very recently the general consensus was that individual differences in cortical organization made any purely
anatomically based definitions of visual areas highly unreliable, even in primary visual cortex.¹

The non-invasive delineation of human visual cortical areas in vivo has brought forth immense amounts of research elucidating many functional properties of the human visual system. As impressive as these achievements are, vision based fMRI research is still a new and developing area, which means there are many unexplored or insufficiently explored topics we are interested in examining. Specifically, fMRI research concerning the functional and organizational properties underlying various visual impairments has started to receive much more attention in recent years, but many unanswered questions remain within this domain. It should be emphasized that there are undeniable methodological shortcomings in many fMRI studies of visual impairments (Binda et al., 2013; Wandell, Dumoulin & Brewer, 2007; Vanni, Henriksson & James, 2005). Our intention here is to optimize retinotopic mapping methods in order to most reliably and efficiently investigate the organization of early visual cortex on both healthy and visually impaired populations. All considerations discussed here have been made with these goals in mind.

*Phase-encoded or traveling wave retinotopy*

The phase-encoded method (often referred to as the traveling wave method) was the first widely accepted fMRI retinotopic mapping paradigm (Engel et. al., 1994; Sereno et. al., 1995) and is still the standard choice for many researchers. This method uses two

¹While many researchers still hold this view (Brewer & Barton, 2012), recent work has shown convincing evidence for reliable and consistent anatomically based retinotopic mapping approaches (Benson et al., 2012). This will be discussed more in the future directions section.
stimuli shown separately: a rotating wedge that establishes the polar angle of visual field along with an expanding and contracting ring that establishes the eccentricity of visual field. The stimuli in this method (and essentially all low level visual area retinotopic mapping methods) are composed of high contrast, flickering checkerboard patterns that are highly efficacious in activating low level visual areas with the added benefit of circumventing unnecessary stimulation in regions that are not retinotopically organized (Thirion et. al., 2006, Brewer & Barton, 2012).

Since the conception of phase-encoded retinotopy there have been a variety of tweaks and adjustments implemented in the design of the stimuli, however, the general principles of the procedure remain consistent across these small changes. Some researchers use stimuli that move in discrete steps along visual field (Smith et. al., 2001) while others use continuously moving stimuli (Engel et. al., 1994; Sereno et. al., 1995). Several attempts at increasing efficiency have also been implemented such as using two wedges that symmetrically stimulate cortex in the left and right hemispheres (Slotnik & Yantis, 2003) and using two expanding or contracting rings simultaneously (Tootell et. al., 1998). Regardless of minor protocol adjustments, the requisite principle of phase-encoded retinotopy is the fact that the stimuli travel along the visual field at a known frequency, thus providing temporal information that can be analyzed in conjunction with fMRI measurements of cortical activity.
Fig 1.1 Phase-encoded stimuli A.) An example of a rotating wedge along with B.) an expanding and contracting ring. The wedge stimulus is typically shown rotating twice, once clockwise and again counterclockwise. Similarly, the ring is presented as expanding and contracting separately. The mean phase delays at each voxel can be determined by taking into account both directions each stimulus cycles through visual field, effectively canceling any residual phase delays (Sereno et. al. 1995).

Specifically, the analysis is intended to find the best fitting harmonic to the measured fMRI time series of each voxel in order to establish the stimulus location that is the most effective at activating that point of visual cortex (Engel, Glover and Wandell, 1997). This is accomplished by performing either a cross-correlation analysis or Fourier transform on the normalized fMRI time series data (Hansen, David and Gallant, 2004; Sereno et. al., 1995). From these computations of signal coherence retinotopic maps can be examined. A standard procedure for doing this involves segmentation and unfolding of the 3-D functional and anatomical data to form a surface based visualization, where the phase reversal in polar angle representations can serve as the basis for delineating early visual areas (Brewer et. al, 2005; Brewer & Barton, 2012).

For a significant period of time, the aforementioned delineation of visual areas through the phase-encoded retinotopic mapping method was without question the most efficient and reliable paradigm available. While other techniques were shown to be
useful as rudimentary alternatives for defining early visual areas (Fox et. al, 1987; Schneider, Noll & Cohen, 1993), these designs provide less precise measurements of visual area boundaries than are necessary for in-depth examinations of retinotopic organization and do not provide any information regarding the spatial tuning properties within visual areas (Dobre et. al., 2001; Wandell, Brewer, & Dougherty, 2005). Researchers yearned to improve phase-encoded retinotopic mapping techniques and subsequently built upon the aforementioned principles with the population receptive field (pRF) paradigm (Dumoulin & Wandell, 2008).

Population Receptive Field Estimation (pRF)

A method that is directly inspired from phase-encoded retinotopy, pRF has been shown to surpass previously used techniques in terms of accurate visual area delineation as well as providing quantitative measures of individual voxel receptive fields throughout visual cortex. Namely, the pRF estimation paradigm incorporates the phase-encoded, expanding and contracting rings along with rotating wedges. The important additions initially developed by Dumoulin and Wandell (2008) in the pRF retinotopic mapping technique are the inclusion of mean luminance periods (just a blank screen with no flickering checkerboard stimuli) and moving bars. These moving bars are not considered to be phase-encoded like the wedge and ring stimuli because each of the four bars have different orientations and each bar will move in opposite directions across visual field, never being repeated within a single run. The mean luminance period is intended to serve as baseline condition and helps estimate receptive fields in higher level visual areas with large receptive fields that respond to such a considerable portion of visual field that
detecting the spatial tuning of voxels is difficult, such as in lateral occipital cortex (LOC). While the mean luminance periods and the moving bars are important additions, the most significant methodological improvements are the computational approaches intended to model neuronal population receptive fields.

Before delving into this quantitative framework of pRF modeling it is important to explain the logical basis of population receptive fields, since by principle they are vastly different than the traditional definition of single cell receptive fields. Because there is a significant limitation of spatial resolution in fMRI measurements compared to single cell electrode recordings and even extracellular recordings, receptive field estimates have to be made using populations of neurons contained in a single voxel. This may appear to be a considerable methodological hindrance, but it is argued to be a reliable work-around for estimating receptive fields at a larger scale (Victor et. al., 1994; Dumoulin & Wandell, 2008). Besides making intuitive sense that nearby neurons are going to have similar receptive fields, it is a well substantiated assumption that local average measurements (the averaging of all neuronal activity within a single voxel) exhibit distinguishable spatial tuning properties² (Smith et. al., 2001; Thirion et. al., 2006; Kay et. al., 2008) that provide the information necessary to estimate the location and size of receptive fields for individual voxels. Neurophysiological observations at a millimeter scale also support the viability of fMRI population receptive field estimates, albeit indirectly. Sereno et. al. (1994) detail their consistent ability to find intermediate

² Tuning properties for orientation (Boynton & Finney, 2003) and spatial frequency (Singh, Smith, & Greenlee, 2000) at the scale of voxel sized measurements are two additional examples that support the characterization of receptive fields based on large neuronal populations.
receptive fields in between two single cell electrode recordings of neurons located in the visual cortex of owl monkeys and also point out that receptive field centers rarely deviated by a significant amount when they gradually sampled extrastriate neurons approximately 50 μm apart from one another.

These population receptive fields are estimated in Dumoulin and Wandell’s (2008) method by fitting a two dimensional Gaussian model of the averaged neuronal population within each voxel using the formula

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

Each modeled pRF estimates the three parameters \(\sigma, x_0\) and \(y_0\) independently for every voxel, where the standard deviation \(\sigma\) is the Gaussian spread of the receptive field centered at the Cartesian coordinates\((x_0, y_0)\). Using the effective stimulus defined by \(s(x, y, t)\) every model pRF is used to calculate a predicted pRF response

\[
r(t) = \sum_{x, y} s(x, y, t) g(x, y)
\]

Convolving a model hemodynamic response function \(h(t)\) with the predicted response \(r(t)\),

\[
p(t) = h(t) * r(t)
\]

the optimal pRF parameters can be found by minimizing the residual sum of squares (RSS) between the actual measurements \(y(t)\) and the above time series prediction \(p(t)\) with a scaling factor \(\beta\)

\[
RSS = \sum_{t} (y(t) - p(t)\beta)^2
\]
This is all accomplished in a “two-stage coarse-to-fine search” (Dumoulin & Wandell, 2008). The coarse fitting stage involves the generation of approximately 100,000 time series predictions with differing pRF model parameters $\sigma, x_0,$ and $y_0$ that essentially constrain the resulting pRF estimates to plausible sizes and locations on a two-dimensional grid that is representative of the visual field. Next, for every voxel estimated as having a significant amount of variance explained in the first step, a finer grid search is performed and the final pRF location and size estimates are determined from these results. It should be noted that these pRF estimates are constrained to not significantly differ from the estimates determined in the coarse fitting stage.

This is just one example of how researchers can implement the pRF paradigm to obtain receptive field estimates. Since the original fMRI pRF publication (Dumoulin and Wandell, 2008) a variety of other Gaussian fitting protocols have been devised, including oval Gaussian models and multiple Gaussian pRF models (Winawer et. al., 2010). Other additions to the quantitative framework for pRF estimation include summary results of pRF laterality (Amano, Wandell & Dumoulin, 2009), functional connectivity models of pRFs across visual areas (Haak et. al., 2012) and perhaps most significantly, the implementation of a static power-law nonlinearity following the initial 2-D Gaussian fit that amends the initial assumption of the pRF model that contrast patterns would linearly sum together through visual field, known as the compressive summation model (Kay et. al. 2013).

It is worth emphasizing that the pRF estimation has become the paradigm of choice not only for researchers interested in studying healthy retinotopic organization of
human visual cortex, but also for investigating the perceptual implications arising from a variety of visual impairments. These impairments include achiasma (Hoffman et. al., 2012), achiasma with comorbid nystagmus (Prakash et. al., 2010), central scotomas found in macular degenerative diseases (Baseler et. al. 2011; Haak, Cornelissen & Morland, 2012), Alzheimer’s disease induced visual deficits (Brewer & Barton, 2011), a hemispherectomized (surgically removed brain hemisphere) patient (Haak et. al., 2012), and a blind individual with surgically repaired sight (Levin et. al., 2010). Researchers that advocate using pRF modeling in studies of visually impaired populations point out that this paradigm is “particularly important for the measurement of visual plasticity in humans” (Brewer & Barton, 2012) which is certainly a vital tool for furthering our understanding of many visual impairments. However, there is evidence that the estimates obtained in many of the aforementioned publications may suffer from methodological artifacts (Binda et. al., 2013, Haak et. al., 2012) and that additional model based evidence would be necessary to justify certain claims concerning radically dynamic and brief reorganization of receptive fields reported in several studies (Wandell & Smirnarkis, 2009). *(see the discussion section for further elaboration concerning these potential biases)* Nonetheless, the pRF method of retinotopic mapping is a widely popular and powerful paradigm for researchers studying the cortical organization and visual processing of both healthy and visually impaired populations.

*Multifocal Retinotopic Mapping*

Multifocal retinotopic mapping using fMRI (mffMRI) actually predates the pRF paradigm by several years, originally shown by Vanni, Henriksson & James (2005) to be
at the very least a viable, qualitative counterpart to phase-encoded retinotopic fMRI mapping measurements. The method acquires its name from the use of multiple, separate, and essentially orthogonal sequences that determine the time course of discreet regions of visual field stimulation. This is significantly different from the phase-encoded and pRF retinotopic mapping techniques, both of which typically present only one, unbroken and coherent stimulus at a time, stimulating visual cortex in a highly temporally correlated manner as it periodically moves across adjacent visual field locations (Sereno et. al., 1995; Engel et. al., 1994). A multifocal stimulus is designed to avoid this large degree of temporal correlation through the use of pseudo-random binary maximum-length sequences (often referred to as m-sequences) (Slotnik et. al., 1999; Vanni, Henriksson & James, 2005). This is accomplished by dividing the visual field into separate regions and presenting the same high-contrast flickering checkerboard patterns used in the previously described techniques according to shifted versions of an initial m-sequence (Carney, Ales & Klein, 2006). These sequences have the important mathematical property of having essentially zero autocorrelation with the original m-sequence (Fig. 2.1) (Sutter, 1992; Victor, 1992). (see Reid, Victor and Shapely, (1997) for a detailed review of higher order m-sequence properties and shift register methods for generating m-sequences) Not only does this avoid the correlated stimulation of visual cortex, a significantly greater amount of cortex can be simultaneously stimulated because m-sequences also have the property of being “on” approximately 50% of the time. If the stimulus is designed so that each area of visual field is scaled according to the cortical magnification factor (CMF) (Duncan & Boynton, 2003) and will therefore stimulate equivalent amounts of cortex,
The autocorrelation of an m-sequence 63 bits in length, defined by the equation $2^6 - 1$. Besides the initial, unshifted m-sequence that is of course going to be correlated with itself (shown at $x=1$) but each of the 63 sequences are unique and can be used as regressors in a GLM orthogonality.

Each of these areas would be active for approximately half of the experiment, assuming that the stimulus covers the full extent of visual field. Comparatively, the wedge and ring used in phase-encoded retinotopic mapping stimulate (at most) approximately 25% of early visual cortex at any given moment. This difference is a very important methodological consideration for optimizing fMRI retinotopic mapping efficiency in terms of experiment length and attainable signal to noise ratios (SNRs) (Buračas & Boynton, 2002; James, Ruseckaite, & Maddess, 2005).

However, not all multifocal retinotopic mapping experiments are designed to maximize the amount of cortical activation achievable with an m-sequence modulated stimulus. In fact, the majority of publications that use multifocal stimuli purposefully
avoid activating such large areas of visual cortex simultaneously. Maximally stimulating visual cortex during the course of a retinotopic mapping experiment was initially the goal of most researchers (Slotnick et. al., 1999; Vanni, Henriksson, & James, 2005) since a largely appealing feature of the multifocal technique is the fact that more stimulus driven activation will be measured, leading to the assumption that this extra data should increase statistical power. Because each fMRI data acquisition of functional responses must be separated by approximately 2 second intervals in order to obtain a robust signal due to hemodynamic lags and the time it takes for the net magnetization to return to equilibrium following excitation, it reasons that if cortical responses to stimuli can only be measured according to these unavoidable temporal restrictions then it would be irrational to not measure as much stimulus driven activity as possible, rather than additional neuronal noise. However, the capacity to orthogonally stimulate approximately 50% of visual cortex at once is not necessarily as advantageous as it might appear. Numerous accounts have been made about SNR reduction resulting from simultaneous stimulation of adjacent regions of visual field, termed surround suppression (Fig. 2.2 – 2.4) (Stenbacka & Vanni, 2007; Pihlaja et. al. 2008; Henriksson et. al. 2012).
These are four different multifocal stimuli (one frame) used by Stenbacka and Vanni (2007) to retinotopically map the peripheral visual field at drastically larger eccentricities than is typically done. The authors state that they were only aware of two other groups (Pitzalis et. al., 2006; Tootell et. al., 1995) that were able find retinotopic maps as far into the periphery as were seen in this study (40° vertically and 49° horizontally). Notably, they add that their goal in doing this study was to replicate the other group’s results but using a more efficient design. Some of the poorest results were measured using stimulus D, which the authors argue stemmed from simultaneous activation of bordering regions of the stimulus. It should be mentioned that only the upper visual field portion of these stimuli are shown here, the actual stimuli include mirror presentations in lower visual field.

†“Reprinted from Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology, 118/6, (Stenbacka. L., & Vanni. S.), fMRI of peripheral visual field representation./ 1303-1314., © (2007), with permission from Elsevier.”
Fig 2.3 (A. – D.) Another four versions of multifocal stimuli (Henriksson et. al., 2012) used in arguably the most successful attempt at retinotopically mapping areas further down the visual stream than V1 using a multifocal design. Two important features of these stimuli account for this success; (A) controlling the stimulation to occur only in non-adjacent regions, (B) the use of only 24 discreet regions, and (C and D) implementing object stimuli to measure retinotopic maps beyond V3. The authors compare the stimulus design and presentation used in this study (A) with their earlier multifocal stimulus (E) that failed to produce retinotopic maps beyond V1 (Vanni, Henriksson, & James, 2005). Evaluations of these results with other retinotopic mapping techniques, including pRF (Dumoulin & Wandell, 2008) and Freesurfer’s probabilistic atlas of anatomical landmarks used by Benson et. al. (2012) showed strong support for the methods detailed by these authors.

‡ “Reprinted from NeuroImage, 27(1), (Vanni, S., Henriksson, L., & James, A. C.) Multifocal fMRI mapping of visual cortical areas./ 95–105., © (2005), with permission
Fig. 2.4 Recreations of the stimuli used in Pihlja et. al. (2008) that illustrated how multifocal stimuli with adjacent region activation can result in severe signal attenuation resulting from surround suppression. The authors quantified the amount of signal attenuation seen in the center region (A) resulting from simultaneous activation of bordering regions (B) and cornering regions (C). Bordering regions showed the most suppression on the center region and the cornering regions showed a fair amount of suppression. This explains the inability for multifocal retinotopic mapping methods to capture local activations in areas beyond V1.

This insight should not be taken lightly, because despite the limited amount of multifocal retinotopy studies there have in fact been quite a variety of protocol adjustments implemented into these paradigms and as such, a variety of observations have been made about the efficacies of these designs. Some of these different stimuli are displayed above but there are several other examples, such as one multifocal EEG experiment that broke the visual field into 192 discreet patch locations (Carney, Ales, & Klein, 2006), a 33 section stimulus where the odd number comes from a centered, circular region designated for para-foveal stimulation (Kay et. al. 2008), and a 48 section stimulus spread across visual space that never simultaneously, activates adjacent regions (Butt et. al. 2011). While there is no clear exemplar of the optimal multifocal stimulus, there is overwhelming evidence from these studies and others that support the notion that
there are detrimental effects on SNR that result from concurrent activations of adjacent regions in multifocal retinotopic mapping designs and that this surround suppression significantly affects the detectability of local activations beyond V1.

While there are some concerns regarding surround suppression, a notable advantage of multifocal mapping paradigms is the statistical approach in which the data can be analyzed. Because the stimulus presentation for each individual region of visual field is designed to have essentially zero temporal autocorrelation, multiple linear regression in the form of a typical general linear model (GLM)

\[ y = X\beta + \epsilon \]

can be used to calculate the beta coefficients (effects of interest), \( \beta \), from the design matrix \( X \) of predicted responses, measured neural activations \( y \), and residual noise \( \epsilon \) (James, 2003; Henrikkson et. al. 2012). One particularly important benefit imparted by a GLM analysis is that it provides a statistical designation for the center of mass of cortical activations corresponding to a specific region in visual field (Slotnik et. al., 1999). This local information is not easily interpreted using temporally correlated phase data because in order to analyze these measurements a Fourier transform or cross-correlation analysis must be performed which will inherently blur the results, making it difficult to attribute stimulus driven cortical activity to precise regions of visual field (Vanni, Henrikkson & James, 2005). In addition, because the oscillatory information produced by the phase-encoded stimuli not only fails to provide directly quantifiable local responses of visual field stimulation, there is no way to quantify BOLD responses arising from specific visual field locations and other non-linear influences such as expectation or attention.
(Masuda et. al., 2008). Analyzing these and other factors would not only improve the spatial resolution of retinotopic maps themselves but could also help detect biases in the results (Binda et. al. 2013). There are advantages realized by employing multifocal fMRI (mffMRI) retinotopic mapping methods, both theoretically (Buracas & Boynton, 2002; Kay, Dumoulin, & Wandell 2007) and experimentally (Hendriksson et. al. 2012; Binda et. al. 2013) as well as shortcomings (Pihja et. al. 2008; Butt et. al. 2011), but when used properly the method has proven to be a viable retinotopic mapping paradigm for examining the implications of cortical organization and functional processes underlying normal vision and a variety of visual impairments.

**Results**

We sought to test multifocal and phase-encoded retinotopic mapping methodologies in terms of their ability to most efficaciously obtain retinotopic maps through human visual cortex based on the previously discussed evidence. The initial hypotheses of this project examined here include:

*A. Multifocal retinotopy will be more efficient than phase-encoded retinotopy in terms of achieved signal to noise ratios at similar scan durations.*

*B. Multifocal retinotopy will be more efficient than phase-encoded retinotopy in terms of achieved signal to noise ratios using shorter scan durations.*

*C. Multifocal retinotopy will produce clearly delineable visual areas V1, V2, V3 and hV4 that are qualitatively comparable to phase-encoded retinotopy.*
D. Multifocal retinotopy will produce maps with finer spatial resolution of corresponding visual field locations in early visual areas than phase-encoded retinotopy.

**Simulation of phase-encoded and multifocal retinotopic mapping methods**

Prior to scanning subjects a simulation was implemented that included a variety of settings such as block length, TR, noise level, and experiment length for multifocal and phase-encoded retinotopic mapping designs, primarily intended to be a proactive way to determine the optimal protocol for maximizing SNR and minimizing scan time in actual experiments. An added advantage the simulation provided was a means to quantify the predicted efficiency of both retinotopy methods across numerous mixtures of the aforementioned settings, so that optimal methodological adjustments based on these results could be decided without having to perform time intensive fMRI experiments. All parts of the simulation were programmed using MATLAB® (2012a, The MathWorks, Natick M.A.) software.

Simulating the two retinotopy experiments can be broken down into the following steps: creating the appropriate stimulus sequences, designing a template brain with reasonably realistic spatial relationships, convolving the stimulus sequence with a canonical hemodynamic response function (HRF) on the template brain to simulate neural activity, and then analyzing this activity to provide meaningful SNR results. The precise techniques used to quantify SNR are detailed in the appendix, rather than here, because the computations comprise several additional steps when done with real fMRI data that are not necessary in the simulation. All other characteristics of the simulation
described in this section share the same essential principles as the methods used for the fMRI experiments.

To provide a meaningful context behind how each step was executed it is first necessary to explain how some of the settings determine the overarching design of the simulation as it relates to the temporal and spatial dimensions of real life fMRI data acquisition. For the purposes of the simulation, only one dimension was necessary to establish cortical space, so each voxel of the simulated cortex was represented as a single column of a two dimensional grid. The rows of this grid represented each simulated acquisition throughout the time course of the experiment. It should be emphasized that each individual voxel is modeled to be tuned to a single, preferred visual field location in which the stimulus would be presented during an actual experiment. To quickly illustrate how this grid simulates a simplified fMRI retinotopic mapping experiment, consider a scanning session that lasts 13 minutes (780 seconds) where the TR is set as 2 seconds and the acquired volume constitutes only 74 total voxels. The simulated data would simply be a 390x74 matrix, signifying that there were 390 acquisitions (each lasting 2 seconds) during a 13 minute scan of a brain area the size of 74 voxels.

Within the basic framework of the simulation determined by the experiment length and total number of voxels, the stimulus sequences are constructed. The process is of course different for the multifocal and phase-encoded stimuli. For the multifocal simulation, the generation of the initial m-sequence was accomplished using Buracas and Boynton’s (2002) shift register function where the only necessary input is the power variable, x, in the equation $2^x - 1$ that defines the length of the m-sequence. Next, for
each voxel the m-sequence was shifted to guarantee that each ensuing sequence would be orthogonal to one another. Since m-sequences are inherently binary, a one was placed at every time point where a given voxel is in the “on” condition (meaning the visual field location corresponding to that voxel is stimulated), otherwise a zero was placed. For the phase-encoded simulation, the stimulus sequence was designed to model the periodic cycling of a wedge and a ring through visual field during separate runs. The separate runs were subsequently concatenated to form one continuous time course as it is common practice to break up the wedge and ring portions of a phase-encoded retinotopic mapping design and also mirrors the procedure described in the fMRI experiment results. The simulation for both methods includes 12 seconds worth of mean luminance (no stimulus presentation) at the beginning and end as this was the protocol used in the actual fMRI experiments. Also in the phase-encoded simulation mean-luminance periods were inserted between the concatenated runs because these are also present in the actual fMRI experiment. The design matrix of each voxel’s modeled time course used in the phase-encoded simulation is displayed in Fig. 3.1.
The stimulus sequences generated in the previous step are eventually convolved with a canonical HRF (appendix A.1) to simulate the neural processes measured in real fMRI experiments, but prior to that the spatial interactions of nearby areas of cortex are accounted for by applying noise to voxels that border stimulus driven activity. This added noise is weighted to reflect the degree of proximity between the voxel locations. It is difficult to quantify appropriate weights for these spatial interactions, but as long as these values remain consistent across the simulated methods it is reasonable to assume that the relative comparison between the two designs should not vary significantly. The weighted interactions of nearby voxels used in the simulations were applied according to the diagram below (Fig 3.2). This proximity matrix is multiplied with the convolved
neural activity and the SNR is then quantified for each voxel in order to meaningfully compare the efficiency of each stimulus design (see appendix for details). In addition to the noise arising from nearby locations of cortex additional Gaussian noise is imposed on the simulated cortical activity in each simulation by a certain factor, or ‘noise level’.

Several figures are provided that illustrate the results of these simulations with respect to the influence of experiment time and random noise on SNRs.

<table>
<thead>
<tr>
<th>Noiselevel:</th>
<th>0</th>
<th>0.5</th>
<th>1</th>
<th>1.5</th>
<th>2</th>
<th>2.5</th>
<th>3</th>
<th>3.5</th>
<th>4</th>
<th>4.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multifocal:SNR</td>
<td>42.27</td>
<td>37.49</td>
<td>28.03</td>
<td>19.59</td>
<td>13.91</td>
<td>10.29</td>
<td>7.91</td>
<td>6.28</td>
<td>5.14</td>
<td>4.39</td>
</tr>
<tr>
<td>stdev</td>
<td>9.4</td>
<td>7.26</td>
<td>3.97</td>
<td>1.91</td>
<td>0.98</td>
<td>0.51</td>
<td>0.37</td>
<td>0.22</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td>Phase-encoded:SNR</td>
<td>5.72</td>
<td>5.66</td>
<td>5.49</td>
<td>5.21</td>
<td>4.83</td>
<td>4.44</td>
<td>4.09</td>
<td>3.76</td>
<td>3.52</td>
<td>3.29</td>
</tr>
<tr>
<td>stdev</td>
<td>0.29</td>
<td>0.29</td>
<td>0.27</td>
<td>0.22</td>
<td>0.19</td>
<td>0.15</td>
<td>0.16</td>
<td>0.15</td>
<td>0.13</td>
<td>0.15</td>
</tr>
</tbody>
</table>

**Table 3.1** Sample of SNRs and standard deviation for both simulations. The simulated experiments were equivalent to 13 minutes scans each. These values are the average SNRs obtained of all voxels with vary degrees of artificial noise.
Fig 3.2 (A) The weighted proximity matrix representing the spatial relations of neighboring voxels in the simulated cortex. (B) Relative spatial interaction weights for edge-adjacent voxels (light blue) and corner-adjacent voxels (darker-blue) to the upper left corner voxel. (The navy blue represents 0 spatial interaction). Because each voxel is designed to have a preference for a 2-D visual field location but are represented along only 1 dimension in the simulation, it might not be clear which voxels prefer similar polar angles or eccentricities. A subtle clue can be seen along the off diagonal elements where the weights bordering the identity elements (colored red) drop. This corresponds to the weights of corner-adjacent interactions and a change in angle preference. Essentially each grouping between these shifts represents voxels that prefer a similar polar angle.
Multifocal and phase-encoded SNRs by noise level (all voxels)

Fig 3.3 Each simulated voxel (48 total) and it’s SNR for both the multifocal and phase-encoded experiments at different noise levels, but with similar experiment times. (~13 minutes). Much higher SNRs were seen in the multifocal simulation until the system noise reached a level of 6. This can be seen by the crossing in figure 3.4 as well. The phase-encoded SNRs are significantly more resilient to noise. Higher noise levels are continued on the next page.
Fig 3.4 The individual voxels themselves did not appear to vary a great deal under different noise conditions, so average SNRs of all 48 voxels were obtained with noise set to 20 different levels (0 to 9.5, intervals of 0.5). From these results it is apparent that under conditions where relatively small amounts of random noise are present, the multifocal method has an exponentially more robust SNR compared to the phase-encoded method. When using the phase-encoded technique there appears to be limitations on the attainable SNR even when there is little to no random noise, however even in exceedingly noisy environments the phase-encoded SNR does not drop considerably, which reiterates that the signal may not be as susceptible to noise promoting artifacts.
Fig 3.5 The effect of experiment time on the SNRs observed in simulations of our multifocal retinotopic mapping technique. Not surprisingly, SNRs increased linearly with increased block lengths. Important inferences can be ascertained from these slopes concerning the relative effects of experiment length and random noise.
Because the multifocal retinotopy simulation using 12 seconds block lengths and the phase-encoded retinotopy simulation shown in the previous figures represent real experiments that would last an approximately identical amount of time (~13 minutes), the average of all voxel SNRs in both simulated experiments was calculated when varying levels of system noise were imposed on both.

In order to examine the efficiency with regards to the time necessary to obtain these SNRs, the multifocal simulation was performed using different block lengths but with the same 63 bit long m-sequence (Fig. 3.5). Block lengths of 4, 6, 8, 10, 12 and 14 seconds were simulated and would be equivalent to an fMRI scan lasting from a little under 5 minutes to approximately 15 minutes. This was repeated with varying amounts of artificial noise.

Discussion of Simulations

The simulations provide strong evidence that the multifocal stimulus achieves considerably greater SNR than possible using the phase-encoded method, consistently for all voxels (visual field locations) and at much shorter experiment lengths. This is in agreement with the initial hypotheses, A and B (C and D were not testable using only a simulation). Since considerable care was taken to design the simulations in a manner as representative of the actual fMRI experiments as possible, these results suggest that the optimal retinotopic mapping paradigm is the multifocal method, at least according to individual voxel measures of SNR. However, there are several considerations regarding the simulation design and the subsequent results that should be mentioned. A wide range of random noise was included in these simulations and the two methods were affected
very differently by this particular component. What is a realistic level of noise for most fMRI experiments, particularly when studying low-to-mid level vision? In order to get a ballpark estimate of the noise interference, the residuals of several previous retinotopic mapping experiments were assessed. The general observation from the standard deviation of regression residuals was that noise levels typically fell in the range of 1.5 to 2. This should be taken into consideration when interpreting the previously shown simulation results.

Additionally, several assumptions were made when implementing this simulation that might not accurately depict real life fMRI measurements of neural activity. The spatial relations, while similar in the phase-encoded and multifocal simulations, were very simplified depictions of very complex neural interactions. Even if the somewhat arbitrary weightings of neighboring voxel interference are relatively sound, there is no accounting for a variety of other neural connections and hemodynamic implications, including feedback from other visual areas (Angelucci et. al., 2002) broad lateral inhibitory signaling (Deangelis, Freeman, & Ohzawa, 1994) and potential saturation of the BOLD response (Harel et. al. 2002). By simplifying the spatial relations to a fixed weighting scheme, many higher order interactions that encompass the phenomenon of surround modulation commonly cited as a limiting factor in multifocal retinotopic mapping studies (Pihlaja et. al., 2008; Stenbacka & Vanni, 2007) might not be captured in the multifocal simulation, resulting in overestimates of SNRs.

Another assumption of the simulation was that attentional and predictability factors would equally influence the observed cortical activity resulting from both stimuli.
This is an important consideration because of the drastically different stimulus presentations involved. A highly predictable stimulus comprising a small portion of visual field may have a very different global influence on neural activity compared to an unpredictable stimulus that subsumes a vastly larger portion of visual field (Kastner et. al., 1999). It is hard to say exactly how each stimulus would influence a real subject’s attentional processes and the subsequent neural implications, however there is evidence that predictable stimuli show greater signal attenuation than non-random stimuli (Binda et. al., 2013; Musada et. al., 2008) meaning the phase-encoded SNRs are potentially overestimated in this regard. Other factors not captured in the simulation include differences in BOLD response, cortical organization, and other individual variability which would be interesting to implement but do not systematically bias the results.

*Functional MRI Experiment*

*Magnetic Resonance Imaging*

To verify the applicability and real life reproducibility of the simulation fMRI experiments were performed using protocols and stimuli that directly corresponded to the simulation described above. A total of 6 subjects (ages 19-25) with normal or corrected-to-normal vision were scanned, however one subject’s scan was not analyzable because of an inability to accurately co-register their anatomical and functional data. The experiments were approved by the Institutional Review Board of The Ohio State University and all subjects gave their informed consent to participate. It should be pointed out that several modifications were implemented as the project evolved, such as removing dual runs of the same experiment in favor of a single run and tweaking the
any significant differences in protocol will be detailed alongside the corresponding results for clarification. All scans were performed at the Center for Cognitive and Behavioral Brain Imaging (CCBBI) located at The Ohio State University. Images were acquired using the CCBBI’s 3-Tesla Siemens magnetic resonance imaging Total Imaging Matrix (TIM) system with a 12-channel head coil. For every subject, a high-resolution T1 weighted MPRAGE sequence was used to acquire structural data with the parameters of 1x1x1mm voxel size, 1900 ms TR, 4.44 ms TE, lasting 6 minutes 50 seconds. Functional data was obtained using a gradient echo, echo-planar sequence with the parameters of 2x2x2mm voxel size, 2000 ms TR, 28 ms TE, 20 total slices oriented perpendicular to the calcarine fissure, and a 72° flip angle. To assist in the co-registration of the partial coverage functional data to the anatomical data, a whole brain, gradient echo echo-planar sequence consisting of a single acquisition 25 seconds in length was also performed.

**Stimuli**

The stimuli were designed using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) in the MATLAB® programming environment. A circular outline of visual field was divided into 72 regions that measured 8.53° in radius. Each region was formed from the intersection of 12 equally spaced wedges and 6 rings that were scaled according to a cortical magnification factor (Horton & Hoyt, 1991) so they have equal area representations within V1. The midpoint eccentricities of the 6 ring were 0.325°, 0.808°, 1.577°, 2.725°, 4.440°, and 7.000°. For both of the retinotopic mapping paradigms if a region was activated, a contrast reversing 4x4 checkerboard was presented. At the center
0.15° of visual field there was a small dot shown throughout the functional scans in order to guide fixation. Additionally, a fixation task was implemented where the participants were instructed to respond by pressing a button whenever the grey dot turned red for a fraction of a second. These responses were recorded so it could be verified that subjects were properly maintaining fixation throughout the scan.

The multifocal stimulus was symmetrically modulated by identical m-sequences according to hemisphere (not hemifield) (Fig 4.1). By designing the stimulus with this symmetric property only 36 total m-sequences were necessary to control the time course of each region. When assigning a shifted m-sequence to a region, care was taken to separate the location of m-sequences only shifted by one block to ensure they were in non-adjacent locations of visual field. This step was probably unnecessary but was done to prevent predictability effects arising spatially and temporally proximal stimulus presentations, as highly unlikely as this possibility sounds. In total, there were 63 blocks (m-sequence length defined by the equation $2^6 - 1$) lasting 12 seconds each. Including the 12 seconds at the beginning and end of each session where no stimulus was presented, the total run time was 780 seconds (13 minutes) in which 390 acquisitions were performed.
Fig 4.1 A.) The empty, discreet stimulus regions defining visual field used in the fMRI experiments. B.) For the mffMRI runs of the experiment there are a total of 36 regions symmetrically assigned according to hemisphere location (see labeling in A.) and modulated by identical m-sequences (72 total regions). The inner most ring follows the same labeling scheme as the other rings (11-16).

The phase-encoded stimuli included two rotating wedges that symmetrically stimulated cortex in the left and right hemispheres analogous to the symmetrical modulation of the multifocal stimulus, along with an expanding ring presented in a separate run. Both phase-encoded stimuli moved in discreet steps through visual field, bounded by the same 72 regions defined in the multifocal stimulus arrangement (Fig. 4.1A), in 4 second blocks. Each of the rotating wedges returned to its starting location every 48 seconds, but because they are symmetrical the wedges cycle at a period of 24 seconds which is repeated 10 times. The run time of the rotating wedge portion of this experiment ended up being 8.4 minutes (504 seconds) during which a total of 252 data acquisitions were performed at a repetition time of 2 seconds. The expanding ring had a
period of 24 seconds and cycled through visual field 9 times. The expanding ring runs lasted 4 minutes (240 seconds) during which 120 acquisitions were made.

Preprocessing

A standard preprocessing pipeline was used for all data included in the analysis. Using the fMRI analysis package AFNI (Cox, 1996), motion correction and co-registration of functional data to the anatomical MPRAGE were performed. The results were specifically analyzed without any spatial or temporal smoothing applied to the functional data, but the wedge scans used to delineate visual area boundaries were smoothed at 2mm, 4mm, 6mm and 8mm FWHM (full width half maximum) for improved visualization on the cortical surface. Prior to any analysis the functional data was also normalized to percent signal change.

Defining ROIs

Numerous attempts were made to delineate the boundaries of V1, V2, V3 and hV4 using only the data obtained using the multifocal method. Despite the expectation that orthogonal activations of regions along the horizontal and vertical meridians, whether they occur simultaneously or not, should provide the information necessary to distinguish early visual area boundaries using a general linear contrast, these efforts were unsuccessful. In place of the intended multifocal based ROI definitions, data obtained from in the rotating wedge scans were used to delineate V1, V2, V3 and hV4. For V1, V2 and V3 this was done through a general linear contrast of neural activation at times the wedges bordered the horizontal vs. vertical meridian and for hV4 the contrast was performed using neural activity during times the wedges were purely represented in upper
vs. lower visual field and had no adjacency to meridians (regions labeled with a 2 and 5 in the ones column, Fig 4.1). Flattened cortical surface maps were generated using Freesurfer’s built-in volumetric segmentation routine (Dale, Fischl, & Sereno, 1999) which were used to visualize the meridian contrast results in the neuroimaging software SUMA (Saad & Reynolds, 2011). The borders of visual areas V1, V2, and V3 were manually defined by the shifts between horizontal and vertical meridian representation along the cortical surface starting at the occipital pole. Area hV4 was delineated using the same procedure, except that the upper vs. lower visual field contrast was used to manually define the ventral boundary.

Results

To identify the extent that local neural activity corresponding to an individual region of visual field could be determined, SNRs were calculated and mapped onto cortex (see appendix A.2 for details concerning the methods). First, to ensure the test-retest reliability of the multifocal method, an intra-subject analysis of 2 separate runs separated by approximately 30 minutes was performed (Fig. 4.2). The later run showed a slight attenuation in SNR at each eccentricity, but this was not significant. Furthermore the overall curve illustrating the relative SNRs in different eccentricities was remarkably similar across runs, suggesting that the lower SNRs observed in the second run resulted from global influences such as scanner noise or participant fatigue, rather than variance in local neural activity. Another step vital in verifying the reliability of the multifocal design was to compare the results of this exploratory technique to the well-established
Test-retest reliability

**Fig 4.2** A single subject’s results across 2 runs of the multifocal retinotopic mapping experiment. The SNRs for each eccentricity were averaged for the sake of visualization.

Phase-encoded retinotopic mapping method. Maps of eccentricity and polar angle were created using the peak SNR clusters obtained from 2 multifocal runs and compared to the phase delay results of a cross-correlation analysis using 2 phase-encoded runs of the same subject (Fig. 4.3). The multifocal retinotopic maps were made using discrete values associated with the SNR cluster analysis which had two main consequences: first, there are only 6 different values (colors) seen in each map and second, there are a few locations not ascribed a preferred polar angle or eccentricity. Without any interpolation of the eccentricity and polar angles values or smoothing of the data along the cortical surface the aforementioned outcomes were inevitable. The procedure for creating retinotopic
maps using phase-encoded stimuli integrates both of these processes which is how the continuous spectrum of colors and smooth transitions along the cortical surface emanate. Taking this into consideration, the multifocal maps are qualitatively comparable to the maps derived from the phase-encoded stimuli. However, in the areas past V1 this qualitative resemblance becomes less robust.

The noticeable decline in the quality of multifocal maps in these areas warranted some investigation into whether the SNR reduced accordingly, or if the clusters identified in the analysis simply decreased in size. An analysis of the average SNR determined for each of the 6 eccentricities across the 4 ROIs showed that both of these factors were implicated in the quality deterioration of the multifocal retinotopic maps past V1 (Fig 4.4).

A unique tool afforded by the multifocal technique is that receptive fields can be characterized for each individual voxel, a methodological feature also present in pRF estimation procedure (Dumoulin & Wandell, 2008). To explore this, receptive fields were estimated (see appendix A.3 for an explanation of this process) for a single subject in visual area V1 using the multifocal data and a slightly modified version of the pRF technique described in the introduction (Fig. 4.5 and Fig. 4.6).
Fig 4.3 Eccentricity (A) and polar angle (B) maps in visual areas V1-hV4. The multifocal maps were made from discrete clusters of peak activities driven by stimulation from individual regions of visual field (*described in appendix A.2*).
Fig 4.4 Average SNR of each eccentricity in each of the 4 ROIs. As visual information is processed along the early visual stream there is either an ostensible influx of noise or an inability of the multifocal stimulus to generate a locally measurable signal. Cluster sizes also decreased (not pictured); however there was a great deal of variance in every ROI in terms of cluster sizes, especially at low eccentricities. It should be noted that even the lowest averaged SNRs recovered in hV4 are at least equivalent or greater than the highest simulated phase-encoded SNRs with 0 system noise.
Receptive Fields of 100 Random Voxels in V1

Red = voxels in right hemisphere
Yellow = voxels left hemisphere

Fig 4.5 One-Hundred randomly selected receptive fields estimated in V1
Fig 4.6 All V1 voxels in which a local minimum was found (~1000).
To examine the spatial resolution in which the visual field can be mapped onto cortex (hypothesis 3) individual peak clusters corresponding to a particular region of visual field were mapped onto V1 (Fig. 4.7). The progression of regions with the greatest eccentricity was coherent (in that they follow a logical relative location to one another) and was consistent with the known local cortical representations shown in neurophysiological measurements and other fMRI experiments.
Fig 4.7 Local cortical representations of the visual field regions according to color. One observation not captured above is that most of the small inconsistencies (such as the blue cluster seen on the upper bank of the right hemisphere and the small red cluster on the lower bank of the left hemisphere) are actually artifacts arising from the opposite hemisphere that likely occurred during the inflating/flattening stage of the gray matter segmentation. When viewed in 3D, there is a clear overlap of voxels between the two hemispheres that were captured in the cluster analysis. This demonstration of the fine spatial resolution achievable with the multifocal design would be even better had these artifacts been accounted for at some point.
Discussion

A great deal of valuable information was learned from these fMRI experiments. Retinotopic maps with great spatial resolution can be created from the multifocal stimulus used in these experiments (at least in V1), which supports hypothesis C. Additionally, SNRs similar to the simulation were achievable, but once again this mostly applies to V1. Robust SNRs were still observed in later visual areas; however the quality of the retinotopic maps declined considerably. One result briefly mentioned in the description of the methods was that ROIs could not be delineated using the multifocal stimulus, which opposes hypothesis C. The inability to reliably define visual areas with a multifocal retinotopic mapping design has been noted several times by other groups (Vanni, Henrikkson, & James, 2005; Binda et. al., 2013) which could be a considerable methodological limitation of multifocal stimuli. One group did successfully employ a multifocal design to delineate V1, V2, V3 and hV4 using an effective weighting scheme and a very different stimulus than presented here (Henrikkson et. al. 2012).

The quantification of SNRs of the phase-encoded experiments proved to be exceptionally challenging due to the unforeseen inability to find intersecting phase information for many regions of visual field. Numerous attempts at finding the appropriate intersections of voxels with particular polar angle and eccentricity preferences were unsuccessful, leaving many gaps in the data. It should be noted that when sizeable clusters were actually found in the phase-encoded data that the SNRs were considerably less robust than the multifocal SNRs. This assessment should be taken very lightly though, since a more concrete and reliable examination of this claim would be
necessary to substantiate this mostly qualitative substantiation. This lack of interpretability explains the absence of directly comparable SNRs in the fMRI data.

Future Directions

Retinotopic mapping using fMRI has become an increasingly popular discipline in recent years. There has been a tremendous influx of research pertaining to retinotopic mapping experiment designs and analytical techniques that will continue to improve our understanding. Assessing the efficiency, usability, and potential of all the advancements this field has seen lately is a difficult endeavor and should not be taken lightly. Only a few major paradigms were discussed and investigated here, but considering the findings in recent publications, the results of extensive simulations, and the outcomes of our fMRI experiments, there are some clear recommendations for optimizing retinotopic mapping that should be discussed.

The overwhelmingly consistent observation reported using multifocal designs is the presence of surround suppression arising from synchronous stimulation of adjacent regions of visual field (Pihlaja et. al. 2008; Stenbacka & Vanni, 2007). The multifocal stimulus as it was presented in the aforementioned fMRI study undoubtedly succumbs to this issue. Spatial resolution and SNR are resoundingly good, but past V1 these attributes deteriorate. Most likely this is also the reason why the process of delineating visual area boundaries is difficult, if not impossible using the multifocal stimulus. Desynchronizing the activation of adjacent visual regions will be necessary for thoroughly analyzing local retinotopic organization of areas beyond V1. While this may result in the need for longer
experiments, it is a necessary modification if areas beyond V1 are to be reliably examined.

However, despite these concerns, there is the potential to advantageously utilize the effects of surround suppression if it could be reliably quantified. Recent work has investigated this, although further work should be done (Sharifian, Nurminen, & Vanni, 2013). Approaches such as this would be highly beneficial as it would not only allow for the mapping of local visual field representations in cortex but may also provide information regarding functional connectivity. A great deal can be elucidated by analyzing the functional connectivity patterns within and across visual areas. There have been recent advancements with regard to retinotopic based functional connectivity methods (Haak et. al., 2012). These methods could provide invaluable complementary substantiations regarding the functional mechanisms underlying visual impairments (Bridge, 2011).

Another retinotopic mapping approach that has recently come to fruition and could be a highly useful technique to complement functionally based retinotopic mapping methods is the anatomical delineation of early visual areas based on sulcal landmarks (Benson et. al., 2012). Until the past few years, a purely anatomically based procedure for retinotopic mapping has been widely considered to be a futile endeavor. Should the work by Benson et. al. (2012) prove to be a reliable and accurate retinotopic mapping paradigm, this could easily serve as an objective tool for evaluating functional delineations of visual areas. Visual impairments originating from or resulting in cortical
pathologies such as amblyopia could be investigated using this method to see if there are noticeable differences in the size and shape of visual areas implicated in these conditions.

Ultimately, a fully optimized retinotopic mapping approach is going to require a well-rounded coalescence of techniques. As described here, each paradigm to examine the organization of visual cortex has some shortcomings. Phase-encoded designs are fraught with temporal correlation and a lack of local specificity (Engel, 2012). The framework for mapping population receptive field models rely heavily on user defined parameters and highly constrained computations that can bias results in ways inconspicuous to researchers (Binda et. al., 2012). Multifocal methods often result in attenuated neural activity arising from surround suppression and have not yet reached their theoretical potential. Despite the unfavorable aspects of many retinotopic mapping paradigms, their underlying principles can serve to complement and improve the efficiency and applicability of one another.

Many recently derived techniques have the potential to improve retinotopic mapping measurements in terms reliability and precision that should be considered in addition to the established methods. In the time this manuscript was written several advances in this field were published, including the compressive spatial summation technique which has initially yielded considerably better results than the pRF method (Kay et. al. 2013). Innovations such as this and other non-linear tools can greatly improve the measures obtained in this project to this point, as long as justified and meaningful interpretations are applied to these concepts.
Optimizing retinotopic mapping methodologies for the purpose of examining healthy and clinical populations is an enduring, but valuable process, that has recently experienced a great deal of progression. The ultimate goal of this work was not only to improve the procedural aspects of fMRI retinotopic mapping but to use these developments to advance our understanding of the organizational properties of visual cortex and potentially investigating visual impairments in the hope that clinical populations will be better served by the knowledge attained. Much work needs to be done, but this can certainly be achieved by rigorous examinations and a thorough understanding of the principles behind the numerous retinotopic mapping approaches.
Chapter 2: Spatial frequency tuning profiles of category selective visual areas

Introduction

A fundamental goal in the vision sciences is elucidating the nature in which low-level image properties are integrated along the visual stream to become coherent and meaningful representations of the world we see and experience. Everything our visual system processes begins as low-level inputs of various contrasts, orientations, and spatial frequencies. Of course, what we actually perceive when viewing complex imagery is not simply an assortment of these individual properties. Rather, these properties undergo various neural computations that give rise to structured forms that are meaningful and describable. A lot is known about the cortical areas that play a role in this process. As previously mentioned in the general introduction, there are a host of known category selective areas such as scene selective areas PPA (Epstein & Kanwisher, 1998) RSC (O’Craven & Kanwisher, 2000) and OPA (Dilks et. al., 2013). Other category selective areas include LOC, an area selective for objects (Malach et. al., 1995) and FFA, an area selective for faces (Kanwisher, McDermott & Chun, 1997). What is not fully understood is the sensitivity these category selective areas of the brain have for different spatial frequencies.
The spatial frequency tuning properties at early points of visual processing have been comprehensively studied, at the level of single cells using intracellular electrode recordings (De Valois, Albrecht & Thorell, 1982) as well as at the millimeter level using fMRI (Henriksson et. al., 2008). What these and related studies have found is that tuning to higher spatial frequencies is more prevalent in regions and cells representing low eccentricities, or more central parts of the visual field, within primary visual cortex and other early visual areas like V2 and V3. Likewise, more peripheral cells and regions are more strongly tuned to low spatial frequencies (Thibos, Cheney & Walsh, 1987).

Extremely comprehensive profiling of spatial frequency tuning in early visual areas has been accomplished in numerous studies over the years, but in later visual areas this same feat has not been accomplished, at least not to the same extent. Because receptive field sizes grow increasingly large along the visual hierarchy, it becomes difficult to robustly test spatial frequency sensitivities compared with early visual areas. Some notable studies have examined high-level visual areas and their affinities toward different spatial frequencies using fMRI that will be discussed; however it is first necessary to explain some prevalent theories regarding spatial frequency processing of complex imagery.

Two widely popular theories that deal with complex image representation (in this case scene perception) and are highly related are known as ‘scene gist’ and the ‘coarse-to-fine hypothesis’. The origin of these theories in scene perception comes from Schyns and Oliva (1994) where they describe two experiments where low and high spatial frequency scenes were combined into hybrid stimuli and presented to participants. For the first experiment, two presentation times were used, 30 ms and 150 ms. After being
shown the hybrid stimulus, a ‘normal’ (unfiltered) stimulus was presented and subjects responded either ‘yes’ or ‘no’ as to whether the unfiltered image was contained in the previous hybrid image. The main finding was that in the 30 ms condition subjects were much better at matching the low spatial frequency image than the high spatial frequency image to the unfiltered target images. The opposite was found in the 150 ms condition; participants matched the high spatial frequency image to the unfiltered target images with much higher accuracy than low spatial frequency images. For the second experiment two hybrid stimuli were presented in succession for 45 ms each, where the second hybrid was the same as the first hybrid except that the spatial frequencies of the initial hybrid image were reversed. For example if a low spatial frequency city and a high spatial frequency highway made up the first hybrid, the second would consist of the same city and highway, except the city would consist of high spatial frequencies and the highway would consist of low spatial frequencies. Subjects were asked to vocally categorize the image they saw and it was found that the majority of responses followed a coarse-to-fine interpretation of the images (i.e. the subjects reported that they saw the scene initially shown at low spatial frequencies and then at high spatial frequencies). Schyns and Oliva make a convincing argument that low spatial frequencies play a very important role in scene recognition and that low spatial frequencies are processed before high spatial frequencies, which is where the ‘coarse-to-fine hypothesis’ of visual perception comes from. They also add a logical explanation for this finding, noting that low spatial frequencies provide the coarse description of the global attributes present in an image, which would be a sufficient amount of information to recognize the overall ‘gist’ of a
scene. Edges and finer details which are determined by high spatial frequencies are processed afterwards in order to refine the initial coarse estimate of the image into a more detailed representation that includes local properties. These hypotheses have subsequently been built on and are the foundation of a very popular computational model of scene recognition known as the spatial envelope (Oliva & Torralba, 2001).

The principles behind scene gist and the coarse-to-fine hypothesis provide a reasonable starting point in understanding how spatial frequencies are processed behaviorally as well as from a computer vision standpoint. Building on these ideas, a handful of subsequent studies have examined the spatial frequency tuning of mid-to-high level visual areas. Peyrin et al. (2004) found that there was hemispheric specialization for processing low vs. high spatial frequencies in parahippocampal and occipito-temporal cortex. They noted that low spatial frequency scenes activated right occipito-temporal areas significantly more than high spatial frequency scenes. Likewise, high spatial frequency scenes activated certain occipito-temporal areas in the left hemisphere. A follow up study by the same group examined whether there was an interaction of hemisphere with the order in which scenes were presented. They found that occipito-temporal areas in the right hemisphere responded more to a sequence of scenes presented in a coarse-to-fine ordering (low to high spatial frequency) and that occipito-temporal areas in the left hemisphere responded more to scenes presented in a fine-to-coarse ordering (Peyrin et al., 2005). Based on these two studies, the authors concluded that mid-to-high level visual processing of low spatial frequency information predominantly occurs in the right hemisphere and high spatial frequencies are processed predominantly
in the left hemisphere. It should be noted that no functional localizers were performed in either of these studies, but based on the coordinates where the authors reported that these hemispheric differences occurred, there would likely be extensive overlap with PPA had it been functionally defined.

In another study, Rajimehr et. al. (2011) reported that PPA responded preferentially to high spatial frequencies in humans and macaques. They showed this effect using several types of filtered stimuli, including scenes, faces, and even checkerboards. Interestingly, they report that high spatial frequency face stimuli were just as effective in activating PPA as unfiltered scene stimuli. One potential concern with this study worth mentioning is the absence of contrast normalization across the different spatial frequency conditions. Instead, the power spectra was averaged for all face stimuli and all scene stimuli prior to filtering, which was used as the basis in determining the cut-off frequencies for filtering the scene stimuli. Cut-off frequencies were defined so that low spatial frequency scenes and faces had the same percent power, and high spatial frequency scenes and places had the same percent power. This approach resulted in different amounts of contrast in the high and low spatial frequency stimuli. Ultimately, this may not have impacted their conclusions too drastically since it was actually the low spatial frequency scenes that had the highest contrast, which if anything would bias neural responses in favor of these images.

This finding by Rajimehr et. al. (2011) was surprising for several reasons. First, it seems to contradict much of the work concerning ‘scene gist’, the spatial envelope and the coarse-to-fine hypothesis. Each of these models/theories put a great deal of
importance on the notion that global features (determined mostly by low spatial
frequencies) are sufficient for scene recognition. Another piece of indirect evidence that
appears to be in conflict with this finding is the peripheral visual field bias observed in
scene selective areas (Levy et. al., 2001; Hasson et. al., 2002). By association, this would
suggest that PPA should show greater sensitivity to low spatial frequencies based on the
fact that high spatial frequency tuning in early visual areas decreases at larger
eccentricities (Thibos, Cheney & Walsh, 1987). This is likely an oversimplified
generalization since high-level visual areas - especially those beyond occipital cortex - do
not necessarily share the organizational attributes of lower level visual areas, but is still
worth considering.

It should be pointed out that until very recently PPA has been assumed to be a
functionally homogenous area, but mounting evidence is coming out that suggests this
assumption is wrong. Examples include PPA’s overlap of several visual field maps
(Acaro et. al., 2009), disparate functional deficits resulting from lesions of PPA (Epstein,
2008), and a meta-analysis of neuroimaging studies of humans and single neuron studies
of macaques (Sewards, 2011). These findings introduce the possibility that PPA is
composed of multiple sub-regions that are functionally distinct, particularly along the
anterior-posterior axis. Of particular interest is that Rajimehr et. al., (2011) allude to
there being stronger high spatial frequency effects in posterior-lateral PPA, although this
was a qualitative observation and no further analyses of this claim were performed.

Perhaps the most convincing work that challenges the notion that PPA is
functionally homogenous comes from Baldassano et. al., (2013) in which they found
differential functional connectivity along the anterior-posterior axis of PPA to distinct cortical networks. Specifically, posterior PPA has stronger connectivity with occipital areas including LOC and OPA\(^3\), whereas anterior PPA has stronger connectivity with RSC and ventral prefrontal cortex. Alongside the functional connectivity analysis, an experiment was conducted that examined how responsive PPA was to abstract objects and scene stimuli. One notable finding from this experiment was that anterior PPA was considerable less visually responsive to abstract objects and slightly less responsive to scenes than posterior PPA. Based on this finding and the functional connectivity results, the authors hypothesized that posterior PPA processes low-level visual features and is simply more visually responsive, while anterior PPA is more involved in processing navigationally relevant features and the spatial context of scenes. In a separate study, Nasr et. al. (2014) reached the same conclusions about the different functional roles of posterior and anterior PPA.

With this background in mind, the main goal of this study was to examine the spatial frequency sensitivities of 5 main ROIs: PPA, RSC, OPA, LOC and FFA. Another goal was to test the spatial frequency sensitivity of PPA by hemisphere and by anterior/posterior sub-regions because of the evidence suggesting that these areas play different roles in processing low-level visual information. The hypotheses tested here are:

\(^3\)Baldassano et. al. (2013) refer to OPA as transverse occipital sulcus (TOS), however their functional definition of this area is the same as OPA and explicitly make note of this in their publication. For the sake of maintaining continuity, any reference to this study will use ‘OPA’ rather than TOS.
1. PPA as a whole will show no sensitivity to low or high spatial frequencies. Both left and right PPA will also show no sensitivity to low or high spatial frequencies.

2. Anterior PPA will be more sensitive to high spatial frequencies and posterior PPA will be sensitive to low spatial frequencies.

3. Occipital areas OPA and LOC will be more sensitive to low spatial frequencies, as will FFA. RSC will be more sensitive to high spatial frequencies.

Methods

fMRI protocol

A total of 16 subjects (ages 18-26) with normal or corrected-to-normal vision were scanned, however 6 subject’s data was not analyzable because of an inability to accurately localize regions of interest at sufficiently stringent significance cut-offs. All results reported below were obtained from the 10 usable subjects. The experiments were approved by the Institutional Review Board of The Ohio State University and all subjects gave their informed consent to participate. All scans were performed at the Center for Cognitive and Behavioral Brain Imaging (CCBBI) located at The Ohio State University. Images were acquired using the CCBBI’s 3-Tesla Siemens magnetic resonance imaging Total Imaging Matrix (TIM) system with a 32-channel head coil. For every subject, a high-resolution T1 weighted MPRAGE sequence was used to acquire structural data with the parameters of 1x1x1mm voxel size, 1900 ms TR, 4.44 ms TE, lasting 6 minutes 50 seconds.

Functional data was obtained using a gradient echo, echo-planar sequence using slightly different parameters depending on which scan was being run. For all 10 subjects,
each functional run used 2x2x2 mm voxel size with no slice gap, a partial coverage FOV consisting of 31 total slices obliquely oriented to maximize coverage of occipital and occipito-temporal areas of cortex and a 72° flip angle. To assist in the co-registration of the partial coverage functional data to the anatomical data, a whole brain, gradient echo echo-planar sequence consisting of a single acquisition 25 seconds in length and oriented identically to the other functional runs was also performed. For both the main experiment and retinotopic mapping localizer a TR of 2000 ms was used, and for the face-place-object localizer a TR of 2500 was used. A TE of 22 ms was used in all functional runs for four subjects and a TE of 30 ms was used in all functional runs for six subjects.

*Unstructured Stimuli*

One set of stimuli, henceforth termed the “unstructured” stimuli, was designed in order to examine neural responses elicited by different spatial frequencies, while not confounding the stimuli with semantic information. These stimuli, devoid of any coherent structure, were created using a band-pass Butterworth filter at narrowly defined spatial frequency bands centered around 0.75, 1.5, 3 and 6 cycles per degree (cpd) with random phase (Fig. 5.1). A total of 60 different unstructured stimuli at each of the specified spatial frequencies were created and randomly selected at each instance they were displayed in the main experiment. Each set of stimuli (the 60 images at each spatial frequency) were contrast normalized.
In addition to examining the neural responses elicited from different spatial frequencies devoid of structure and semantic information, we also employed grayscale photographs consisting of four different natural scene categories: beaches, forests, cities and highways. Each of the original photographs was low and high pass filtered at 0.75 cpd and 6 cpd, respectively, using a second-order Butterworth filter. Following this frequency filtering, the contrast of each set of images (meaning the original grayscale, the high spatial frequency (HSF) and the low spatial frequency (LSF) image) were jointly root-mean-square normalized so that the three images have equal contrast. This step ensured that for each image, the full spectrum (FS), HSF and LSF versions differed only in terms of spatial frequency information and not luminance or contrast.
Main experiment design

The main experiment employed a block design that consisted of eight runs, each lasting approximately five minutes and thirty seconds. Every run contained a 20 second block for each of the scene categories (four) and image type (three), making a total of 12 blocks of scene stimuli. Four additional blocks, one for each of the unstructured spatial frequency stimuli sets were also included in each run, making the final number of blocks per run 16. The order of the 16 blocks, each consisting of a different stimulus variety, was randomized across runs and counterbalanced across subjects.

Each image was presented in succession for 800 ms separated by a 200 ms inter-stimulus interval consisting of a gray, mean-luminance background. A fixation cross spanning approximately 3 degrees of visual angle was present throughout the entirety of each run. A total of 8 images were shown in each block, concluding with a period of 12 seconds of mean-luminance fixation prior to the start of the next block to allow the hemodynamic response to return to baseline. Participants were instructed before the start of each run to maintain fixation on the center cross and to simply view the scenes and
stimuli displayed. There were no explicit tasks required of the participants throughout the duration of the main experiment besides passively viewing the various scene and unstructured stimuli.

![Sample experiment sequence. Each of the 16 possible sets of stimuli were included in each run in a randomized order.](image)

**Regions of interest (ROI) localization**

Separate from the main experiment, two runs of an independent face-place-object (FPO) localizer were performed. Each run, lasting approximately seven minutes and thirty seconds involved presenting a series of face, scene, object and scrambled object images in which the subjects were instructed perform a one-back task when an image appeared twice in succession via a button response. Scene selective areas (PPA, RSC and OPA) were defined using a linear contrast of neural activity elicited by scenes vs. faces and objects (Epstein & Kanwisher, 1998; O’Craven & Kanwisher, 2000; Dilks et. al., 2013). FFA (a face sensitive area) was defined using a contrast of responses to faces vs. scenes and objects (Kanwisher, McDermott & Chun, 1997) and LOC (an object sensitive
area) was defined using a contrast of responses to objects vs. scrambled objects (Malach et. al., 1995). All data from the FPO localizer scans was smoothed using a Gaussian kernel of 4 mm full-width half-maximum (FWHM). For the six subjects scanned using a TE of 30 ms, ROIs were delineated in the three dimensional volume using AFNI’s clustering tools, according to the aforementioned contrasts. These ROIs were unilaterally defined using a minimum threshold of p < 0.05. Thresholds were made more stringent in cases where clusters of voxels blatantly not belonging to ROI being defined needed to be separated. For the four subjects scanned using a TE of 22 ms, ROIs were delineated on inflated surface volumes that were generated using Freesurfer’s built in volumetric gray-white matter segmentation routine (Dale, Fischl, & Sereno, 1999) using the neuroimaging visualization software SUMA (Saad & Reynolds, 2011). In order to account for both the natural smoothness of the data resulting from the hemodynamic response time course acquired in these scans and the 4 mm FWHM smoothing applied prior to defining these functional ROIs, a minimum cluster size was calculated using AFNI’s built in AlphaSim function for each individual subject to ensure the clusters forming each ROI were unlikely to occur by chance. In all 10 subjects it was possible to define PPA, OPA, FFA and LOC in both hemispheres according to our criteria. RSC was defined in both hemispheres for 9 out of 10 participants; however for one subject (scanned using a TE of 30 ms) RSC was not delineable in the right hemisphere.

For the delineation of PPA into anterior and posterior sub-regions a simple median split of voxels along the A-P axis of PPA was performed separately in each hemisphere (PPA of each individual subject in each individual hemisphere was divided
according to this median split). This ensured that anterior PPA and posterior PPA had roughly an equal number of voxels per subject and hemisphere. A small discrepancy in size between these two sub-regions, where slightly more voxels would be assigned to posterior PPA, resulted from the arbitrary decision to assign voxels located at the median A-P coordinate to posterior PPA. This was chosen over several alternatives, such as performing a random assignment of these voxels to anterior or posterior PPA or leaving these voxels out of the analyses completely. Random assignment of voxels along the median A-P plane to each subregion was not used in order to ensure our definition of anterior and posterior PPA was consistent for all subjects and to also avoid confounding these ROIs with different amounts of voxels located more medially and laterally, since it has been hypothesized that there are distinguishable functional roles of temporal lobe along this orthogonal plane (Nasr, Devaney & Tootell, 2013). All anterior and posterior PPA ROIs were analyzed using the definition described here; i.e. anterior and posterior PPA simply constitute approximately an equal number of voxels lying along the A-P axis of PPA.

Retinotopic mapping

To delineate early visual areas, standard phase-encoded retinotopic mapping scans were performed (Sereno et. al., 1995; Engel et. al., 1994). This encompassed a single eight minute run of a rotating wedge stimulus (traveling both clockwise and counter clockwise in direction) to determine polar angle and a single four minute run of an expanding and contracting concentric ring to determine eccentricity. The wedge and ring stimuli each consisted of high contrast flickering checkerboard patterns. In order to help
ensure subjects maintained fixation, participants were instructed to fixate on a gray dot located in the center of the screen and respond with a button press when the dot turned red (jittered at 5-15 second increments). The wedge run was performed for all 10 subjects and the ring run was performed for 7 of the 10 subjects.

Preprocessing

All data obtained underwent standard preprocessing procedures using the fMRI analysis package AFNI (Cox, 1996). Functional scans were motion corrected and co-registered to the anatomical MPRAGE in a single step in order to eliminate the need for multiple interpolations of the functional data. Because of our interest in exploring functional implications along the anterior-posterior axis of PPA it was necessary that the anatomical scans remained unmoved from the position in which they were originally acquired which is why the functional scans were aligned to the MPRAGE instead of the other way around. For all MVPA analyses, a 2 mm FWHM smoothing kernel was used. All univariate analyses were performed following 4 mm FWHM smoothing. Prior to any analysis all functional data were normalized to percent signal change.

Univariate analyses

Univariate analyses of the main experiment runs were performed on individual subject data using a standard general linear model (GLM) approach. Incorporated into the full model as regressors were each of the four unstructured spatial frequency stimuli (0.75, 1.5, 3 and 6 cpd), the HSF scene stimuli, the LSF scene stimuli and fixation (no stimulus present) time courses. Six motion parameters (provided from the output of the motion correction step of preprocessing) and low frequency scanner drift regressors were
also included in the model. A canonical hemodynamic response function provided by AFNI (see appendix B for specifics regarding this function, Cohen, 1997) was convolved with a square wave form lasting either eight seconds (corresponding to the length of time each series of images was presented within each block) or 12 seconds (corresponding to the fixation period included at the end of each block).

Beta coefficients signifying percent signal change for each voxel in each condition were modeled using this GLM. Individual subject baselines for each ROI were defined by the mean percent signal change during fixation periods of the main experiment. These baseline values were subtracted from the mean percent signal change in the reported ROIs for all of the stimulus conditions.

**Multivariate Analyses (MVPA)**

To examine the influence of spatial frequency information on distributed patterns of neural activity in the ROIs previously described, multi-voxel pattern analyses (MVPA) were also performed. The residuals following a GLM analysis that eliminated all nuisance regressors (e.g. scanner drift and head motion) were the data used in these multivariate statistical analyses. A four second lag (approximating the delay of the hemodynamic response) was applied to these residuals and then calculated for the 512 volumes acquired of each subject while viewing stimuli (8 runs x 16 blocks x 8 images displayed 1 second each / 2 second TR) during the full time-course of the main experiment. Next, a linear support vector machine (SVM) (Chang & Lin, 2011) was trained to assign labels of either the scene category of the stimulus or spatial frequency (binary classification of high vs. low) of the stimulus. Importantly, training was done in
each instance for each subject on only 7 of the 8 main experiment runs and the data from
the left out run was subsequently provided to the SVM classifier which generated
predictions of class labels for each block. This procedure was repeated eight times,
where each of the eight runs was left out once during the training of the classifier (a
leave-one-run-out, LORO, cross validation process).

*Searchlight Analysis*

Using the Searchmight software package in MATLAB (Pereira & Botvinick, 2011) a whole-brain searchlight analysis was performed for each subject. The searchlight
method employed a Gaussian Naïve Bayes pooled classifier that gradually shifted a
5x5x5 mm template seed throughout each individual subject’s cortex. Scene category
decoding accuracies were stored at the center voxel of each template location as it
traversed cortex. No voxels outside the brain were included for this analysis. Upon
completion of the searchlight, decoding accuracy masks for each subject were
transformed into Montreal Neurological Institute (MNI-152) space and smoothed at 2
mm FWHM. Voxel-wise t-tests were performed on these accuracy maps thresholded at p
< 0.05 (uncorrected) to determine if scene category classification accuracy was
significantly above chance. Cluster correction was applied based on the voxel
dependencies arising from the smoothing process (2 mm FWHM on the residuals used
for the searchlight and the 2 mm FWHM smoothing done on the decoding accuracy
masks) and hemodynamic response.
Results

Univariate

In order to measure the overall activation levels elicited by scenes containing only high (≥ 6 cpd) or low (≤ 0.75 cpd) spatial frequencies in each of the category selective ROIs (PPA, RSC, OPA, LOC and FFA) 10 subjects passively viewed frequency filtered and contrast normalized scenes (beaches, forests, highways and cities) while being scanned. Included in each run were blocks of contrast normalized, unfiltered images that contained the full spectrum of spatial frequency information prior to filtering (see Methods for details). In order to see how different spatial frequencies not confounded with semantic information drove these areas, blocks of unstructured stimuli with narrowly defined spatial frequency bands around 0.75, 1.5, 3 and 6 cpd were also included in the experiment. The five category selective ROIs were defined independently using separate localizer scans and a standard GLM analysis was performed to examine the percent signal change generated in each of the experimental conditions. (see Methods for details).

The three scene sensitive areas PPA, RSC and OPA showed a consistent pattern of greater responses to HSF scenes compared to LSF scenes (Fig. 6.1). This was marginally significant in PPA and RSC ($t_{(9)} = 1.89; p = 0.091$ and $t_{(9)} = 1.99; p = 0.077$, respectively) and was significant in OPA ($t_{(9)} = 3.08; p = 0.013$). After applying a Bonferroni correction for multiple comparisons, this effect in OPA is no longer
Fig 6.1 The percent signal change in each of the experimental conditions. The most robust difference in activation levels between the HSF and LSF scenes was observed in OPA. Furthermore, HSF scenes activated OPA more than any of the ROIs tested. Planned comparison t-tests showed three were marginally significant (p < 0.1) and OPA did significantly respond more to HSF scenes than the LSF scenes (p < 0.05). Error bars represent 95 percent confidence intervals from 0 percent signal change.
significant. LOC exhibited a similar, marginally significant sensitivity towards HSF scenes \((t_{(9)} = 2.097; p = 0.065)\) but FFA responded equally to HSF and LSF scenes \((t_{(9)} = 0.59; p = 0.57)\).

Because of the work done by Peyrin et. al., (2004, 2005) regarding the spatial frequency sensitivity differences in parahippocampal and occipito-temporal across hemispheres, planned comparisons of mean activation to LSF vs. HSF scenes within left and right unilateral PPA were done. There was no main effect of spatial frequency or hemisphere \((F_{(35)} = 0.86; p = 0.36\) and \(F_{(35)} = 2.63; p = 0.13\), respectively) and no interaction of hemisphere with spatial frequency \((F_{(35)} = 0.40; p = 0.54)\). Contrary to Peyrin et. al. (2004, 2005), it was found that right PPA responds more to HSF scenes than LSF scenes \((t_{(9)} = 3.74; p = 0.0046)\). There was no significant difference in left PPA activation levels to either of these conditions \((t_{(9)} = 0.29; p = 0.776)\) (Fig. 6.2A). In addition to this hemispherical analysis of PPA, spatial frequency effects were compared in anterior and posterior PPA (see Methods for details on this delineation). Posterior PPA showed significantly higher sensitivity to HSF scenes than LSF scenes \((t_{(9)} = 2.58; p = 0.029)\) but no such effect was found for anterior PPA \((t_{(9)} = 0.27; p = 0.79)\) (Fig 6.2B). There was no interaction of spatial frequency and anterior/posterior PPA \((F_{(35)} = 0.1292; p = 0.72132)\), no main effect for spatial frequency \((F_{(35)} = 1.03; p = 0.316)\) and there was a main effect of PPA sub-region \((F_{(35)} = 4.73; p = 0.0366)\).
Fig. 6.2 Two approaches for dividing PPA, both based on previous findings (Peyrin et al., 2004; Baldassano, Beck & Fei-Fei, 2013) of functional or connective inhomogeneities attributed to PPA. A significant HSF sensitivity over LSF scenes was found for A. right PPA and B. posterior PPA. The complementary area in both cases did not have this difference in SF sensitivity. Error bars represent 95 percent confidence intervals from 0 percent signal change.

Doing high vs. low comparisons within each ROI for the unstructured stimuli (where the two lowest spatial frequencies and two highest spatial frequencies were combined into separate groups) yielded no significant differences (PPA: \( t(9) = 0.82, p = 0.43 \); posterior PPA: \( t(9) = 0.92, p = 0.38 \); anterior PPA: \( t(9) = 0.63, p = 0.55 \); left PPA: \( t(9) = \))
$t_{(9)} = 0.18, p = 0.86$; right PPA: $t_{(9)} = 1.03, p = 0.33$; RSC: $t_{(9)} = 0.53, p = 0.61$; OPA: $t_{(9)} = 1.76, p = 0.11$; LOC: $t_{(9)} = 0.89, p = 0.40$; FFA: $t_{(9)} = 0.46, p = 0.65$). These null findings are likely because these stimuli simply did not drive these areas to a significant extent.

**MVPA**

In addition to examining mean activation levels we are interested in how much information regarding scene category is conveyed by HSF and LSF scenes in scene-selective regions. It has previously been shown that the scene selective areas PPA, RSC and OPA not only show greater relative levels of activation in response to places and buildings compared with faces and objects (Epstein & Kanwisher, 1998; O’Craven & Kanwisher, 2000; Dilks et. al., 2013), but that the distributed patterns of activity elicited from different types of scenes contain category-specific information (Walther, et. al., 2009; Walther et. al., 2011), multivariate analyses aimed at examining the effect of spatial frequency were performed in these regions. Each of the ROIs chosen had robust mean activation levels driven by the LSF scenes that resembled the responses to the full spectrum and HSF scenes much more than the responses to the unstructured stimuli. These multivariate analyses would be able to ascertain whether, and to what extent, category specific information would still be present in the filtered scenes. On top of that, this multivariate approach can provide further information about whether, and to what extent, spatial frequency itself (high vs. low) is discriminable in the patterns of activity measured from each of these ROIs, especially with regards to the unstructured stimuli since the mean activation levels elicited by these stimuli were very low and difficult to
interpret on their own. LOC and FFA were also included in this analysis due to the sensitivities they exhibit in response to complex stimuli, which allude to the possibility that different spatial frequencies contained in complex visual information are processed differently in these areas as well.

The scene stimuli used in the main experiment consisted of four categories (beaches, forests, highways and city streets) that the classifier was trained and subsequently tested on using the LORO procedure described in the methods section. Of particular interest was testing how the spatial frequency content of the low and high pass filtered sets of scenes would impact the classification rates of scene category from the patterns of neural activity measured in the aforementioned ROIs. Illustrated in figure 6.3, there was a very consistent trend in which scene category was not decodable above chance (25%) when only low spatial frequency information was present in the image. On the other hand, the classifier was able to decode the category of the HSF scenes above chance for all ROIs except FFA. Planned comparisons between the LSF and HSF
Fig 6.3 The distributed patterns of activity measured from every ROI that was analyzed provided more category-specific information for HSF scenes than LSF scenes. LSF scenes were decoded at chance performance across all five main ROIs. This effect was most robust for OPA, the area that also decoded scene category with the highest accuracy in the HSF condition. Error bars are 95% confidence intervals from chance (1/4).

conditions showed that scene category classification accuracies were significantly higher for HSF scenes than LSF scenes using the pattern of neural responses measured in RSC \( (t_{(9)} = 2.95; p = 0.0161) \), OPA \( (t_{(9)} = 3.68; p = 0.0051) \) and LOC \( (t_{(9)} = 2.59; p = 0.0292) \). Scene category decoding of HSF scenes over LSF scenes from PPA was marginally significant \( (t_{(9)} = 1.86; p = 0.096) \) and even FFA showed this same trend \( (t_{(9)} = 1.826; p = 0.101) \), with the caveat that neither HSF nor LSF scenes were decodable above chance from FFA. It should be noted that after using a Bonferroni correction for
multiple comparisons the only ROI in which HSF scene classification was still significantly greater than LSF scene classification was OPA. (see Appendix B/Supplementary Tables for raw decoding accuracies). This provides very strong evidence that OPA has an extremely important role in processing high spatial frequency information of scenes.

As stated above, a great deal of evidence points to the possibility that encoding properties of PPA differ across hemisphere as well as along its anterior-posterior axis (Peyrin et. al., 2004; Peyrin et. al., 2005; Baldassano, Beck & Fei-Fei, 2013). The univariate analysis tested these theories and in the left vs. right hemisphere comparison of spatial frequency sensitivity no difference was observed for left PPA, but right PPA responded significantly more to HSF scenes compared to LSF scenes. Likewise, when partitioned into anterior and posterior sub-regions, anterior PPA showed no significant disparity in mean activation elicited by HSF or LSF scenes, but posterior PPA responded significantly more to HSF scenes. While there were no interactions of PPA sub-region (hemispherically or along the anterior-posterior axis) with spatial frequency condition, these findings combined with previous reports advocated that similar comparisons be done using multivariate techniques as well.

HSF scenes were decodable above chance from PPA ($t(9) = 2.67; p = 0.0254$) and from the posterior and anterior sub-regions of PPA ($t(9) = 5.47; p = 0.0004$ and $t(9) = 5.04; p = 0.0007$, respectively). On the other hand, LSF scene category was only decodable above chance in posterior PPA ($t(9) = 2.28; p = 0.0487$). As seen in figure 6.4, there was a significant two-way interaction between spatial frequency and PPA.
Fig 6.4 Classification rates of scene category for PPA, posterior PPA and anterior PPA in each of the three scene spatial frequency conditions. While the trend seen in PPA and the other ROIs where HSF scene categories are classified much better than LSF scene categories for posterior and anterior PPA, there is a two-way interaction of spatial frequency with PPA subregion. Error bars are 95% confidence intervals from chance (1/4).

subregion \((F_{(35)} = 4.553; p = 0.0396)\). This suggests that there is some functional difference across anterior and posterior PPA, even though both regions follow the same
trend where HSF scene categories are classified with higher accuracy than LSF scenes. It should be noted while HSF scene category classification rates from posterior PPA were better than the classification rates of LSF scenes, this was not significantly different ($t_{(9)} = 1.06; p = 0.3167$). Doing this same comparison for anterior PPA classification rates does show a significant difference for HSF scenes and LSF scenes ($t_{(9)} = 3.66; p = 0.0052$) and no difference in posterior PPA.

Examining PPA by hemisphere, HSF scene category was again classified with higher accuracy than LSF scenes from left PPA ($t_{(9)} = 2.61; p = 0.028$) and from right PPA ($t_{(9)} = 4.19; p = 0.0023$). A two way ANOVA revealed that unlike the interaction of spatial frequency with posterior and anterior PPA, there was no such interaction of spatial frequency between hemisphere and spatial frequency ($F_{(35)} = 4.553; p = 0.439$). Notably, LSF scene categories were decodable above chance in left PPA ($t_{(9)} = 2.45; p = 0.038$), making posterior PPA and left PPA the only two ROIs where this occurred.

Because the scene category of unfiltered photographs has been shown to reliably be decodable from many of the ROIs selected (Walther et. al., 2009), scene category was the initial test used in ascertaining the spatial frequency sensitivities of these ROIs. Another question that could be asked is whether spatial frequency itself would be distinguishable using the patterns of activity measured from these ROIs. For this analysis, the same LORO training method was used, except the classifier was asked to determine what the spatial frequency (low vs. high) of the images was (rather than category). This was done for both the unstructured and scene stimuli. The spatial frequency of the scene stimuli were classified above chance from PPA ($t_{(9)} = 4.30; p =$
0.002), OPA ($t_{(9)} = 3.45$; $p = 0.0073$) and LOC ($t_{(9)} = 11.05$; $p < 0.00001$), but not from RSC ($t_{(9)} = 0.63$; $p = 0.54$) or FFA ($t_{(9)} = 1.61$; $p = 0.142$). In order to do a

**Decoding Accuracies - Scene Category**

Fig 6.5 Decoding accuracies for PPA based on hemisphere. LSF scene category was actually decodable above chance from left PPA. Comparing HSF and LSF scene category classification accuracies within hemisphere shows that HSF scenes were classified correctly significantly more often than LSF scenes across both hemispheres. There was no interaction of hemisphere and spatial frequency. *Error bars are 95% confidence intervals from chance (1/4).*

binary classification of high vs. low spatial frequency for the unstructured stimuli, the two lowest spatial frequencies (0.75 and 1.5 cpd) and the two highest spatial frequencies
(3 and 6 cpd) were grouped together. Spatial frequency of the unstructured stimuli was decodable significantly above chance from RSC ($t_{(9)} = 3.42; p = 0.0072$) but was not decodable from the other four main ROIs (PPA: $t_{(9)} = 1.91; p = 0.088$; OPA: $t_{(9)} = 1.25; p = 0.243$; LOC: $t_{(9)} = 1.77; p = 0.111$; FFA: $t_{(9)} = 1.89; p = 0.0912$)

Based on several hypotheses that the posterior region of PPA plays a significant role in processing low-level properties of images and anterior PPA does not (Baldassano et. al., 2013; Nasr et. al., 2014), activity from these sub-regions were also given to the classifier to see whether these areas could discriminate scene and unstructured images on their spatial frequency content alone. The spatial frequency of scene stimuli was decodable above chance from both posterior and anterior PPA ($t_{(9)} = 3.70; p = 0.0049$ and $t_{(9)} = 3.22; p = 0.0104$, respectively). Interestingly, the spatial frequency of the unstructured stimuli was decodable above chance from posterior PPA ($t_{(9)} = 3.84; p = 0.0039$) but not anterior PPA ($t_{(9)} = 0.59; p = 0.57$) (Fig. 6.6).
Fig 6.6 Decoding accuracies of HSF vs. LSF scenes (teal) by spatial frequency and LSF unstructured stimuli (0.75/1.5 cpd) vs. high unstructured stimuli (3/6 cpd) stimuli (magenta) in PPA and anterior/posterior sub-regions of PPA. Not surprisingly, scenes were decodable by spatial frequency significantly above chance throughout PPA and its sub-regions. Noticeably, spatial frequency of the unstructured stimuli was only decodable above chance from posterior PPA (61%).
A whole-brain post hoc examination of areas located outside of the defined ROIs was performed using the Searchlight MATLAB toolbox (Pereira & Botvinick, 2011) (see Methods for details). This analysis was exploratory in nature and not the emphasis of this study, but a few results are worth mentioning. Qualitatively, the decoding accuracies of HSF scene category were above chance for substantially more cortex than the decoding accuracies of LSF scene category. When comparing the searchlight results for the ROIs in MNI space, there were considerably more voxels/clusters found where HSF scene category was decoded above chance than the LSF scenes. This is in line with the results of the ROI multivariate analysis and two examples are provided below (Fig 6.7/6.8). In anterior PPA no LSF accuracy map clusters survived thresholding, but in posterior PPA there were some significant LSF clusters. However, it should be made clear that these maps were generated at a fairly lenient threshold (p < 0.05, uncorrected) and that with any stricter criteria these LSF clusters do not survive.
Fig 6.7 Group surface maps of searchlight decoding accuracies above chance (p < 0.05) for left and right PPA and the posterior and anterior sub-regions. Green = LSF; Red = HSF; Yellow = Overlap.

Fig 6.8 Group surface maps of searchlight decoding accuracies above chance (p < 0.05) for left and right OPA. Green = LSF; Red = HSF; Yellow = Overlap. Interestingly, right OPA was found to have more LSF represented. This is a potential example of hemispherical differences that subsequent experiments may want to analyze further.
Discussion

Univariate, multivariate and searchlight analyses were used to comprehensively examine the spatial frequency tuning profiles of five category selective regions of visual cortex (PPA, RSC, OPA, LOC, and FFA) along with four subdivisions of PPA (left/right PPA and anterior/posterior PPA). Across these areas a surprisingly consistent trend was found in each of the analyses: category selective areas are predominantly sensitive to high spatial frequencies.

For the univariate analysis, HSF scenes significantly activated OPA more than LSF scenes and HSF scenes marginally activated PPA, RSC and LOC more than LSF scenes. Overall the unstructured stimuli did not drive these category selective areas significantly. Examining PPA by hemisphere and by anterior/posterior sub-regions, posterior PPA responded significantly more to HSF scenes than LSF scenes supporting the findings of Rajimehr et. al. (2011). Contrary to Peyrin et. al. (2004, 2005), right PPA was found to respond to HSF scenes more than LSF scenes. There are some differences in their high and low pass filtering techniques that may explain this discrepancy, in particular they used a high pass filter cutoff of 1 cpd. Additionally, their stimulus presentation time was only 100 ms which is significantly shorter than the 800 ms presentation time used here.

The multivariate analyses captured very similar trends seen from the mean activation levels, with a few very important findings worth emphasizing. Perhaps the most intriguing finding is that LSF scenes drive all of the scene selective areas fairly robustly, well beyond the amount the unstructured stimuli do, but category specific
information is lost. Decoding accuracies of LSF scene category was at chance in every ROI except for posterior PPA and left PPA. What might be behind this discrepancy?

Low spatial frequencies provide information about global attributes of scenes such as openness and expansiveness (Park et. al., 2011), characteristics of a scene that appear to be sufficient in activating scene-selective areas more than unstructured stimuli containing similar spatial frequencies (Oliva & Torralba, 2006). The key here is that even though low spatial frequency scenes only contain coarse levels of visual input, there is still relevant semantic information present in the phase of the image, something that is not found in the unstructured stimuli. Likely this is enough information to indicate that the image being viewed is a scene, but this information lacks category specific attributes. It should also be pointed out that behavioral experiments have shown that scene categorization definitely involves processing high spatial frequency attributes of scenes such as curvature and junction properties (Walther & Shen, 2014).

The results of the MVPA also revealed a two-way interaction of spatial frequency and posterior/anterior sub-region of PPA. Previous researchers have hypothesized that posterior PPA likely plays a larger role in processing low-level visual properties of stimuli (Baldassano et. al., 2013; Nasr et. al., 2014). This significant interaction across PPA sub-regions along with the finding that spatial frequency of unstructured, semantic free stimuli was decodable from posterior PPA and not anterior PPA (although there was no interaction here) would seem to be in agreement with that hypothesis. Of course this is only the tip of the iceberg concerning a full understanding of any functional inhomogeneities found in PPA. More work will need to be done to elucidate more
precisely what processes are occurring along the A-P axis of PPA along with determining the nature of this separation. Are functional differences found along a gradient within PPA, or might there be more precise delineations of function subunits? Based on the functional connectivity of PPA (Baldassano et. al., 2013) these functional differences are more likely to occur along a gradient, but as previously stated, more work needs to be done to answer this question.
General Conclusion

While a great deal is understood about the human visual system, there are still many unsolved questions regarding the neural processes that integrate low-level image properties into complex, meaningful representations of the world around us. The studies described here sought to contribute to our current understanding of this enduring matter by examining both the organizational and functional characteristics along the visual system hierarchy.

Specifically, the first study examined the organizational properties of early visual cortex with the intention of optimizing retinotopic mapping paradigms for use in future experiments. First, using a series of simulations, two retinotopic mapping designs (the conventional phase-encoded technique and the multifocal method) were compared in terms of efficiency and the achievable signal-to-noise ratio under various conditions (such as varying degrees of exogenous noise, different block lengths and different experiment lengths). The main result from these simulations showed that as long as there were realistic levels of noise present, the multifocal method would achieve vastly higher signal-to-noise ratios than the phase-encoded design would at similar experiment lengths. This finding, combined with the many advantages that using a multifocal approach theoretically has (such as the ability to use a block design and fit a GLM rather than doing a cross-correlation analysis, being able to stimulate half of visual cortex at once and being able to map localized regions of visual field to visual cortex with a higher
degree of spatial resolution) suggested that the multifocal method would be the optimal choice.

To test this, fMRI versions of the simulations were performed. The results were far less clear than was expected from the simulations. Beyond V1, the retinotopic maps measured using the multifocal method were scattered and irregular compared to the conventional phase-encoded maps. The declining quality of the retinotopic maps was most likely due to surround suppression from neighboring areas of cortex being activated, a problem that got worse as the size of receptive fields grew in later areas. Ultimately, this was a valuable finding that can help guide the development of retinotopic mapping paradigms in the future.

The second study examined the spatial frequency tuning profiles of category selective areas, PPA, OPA, RSC, FFA and LOC. In addition to these areas, anterior and posterior sub-regions of PPA were delineated and included as ROIs because of several recent reports that suggest PPA may not be a functionally homogenous area. A univariate analysis was performed in order to test for differences in mean activation levels elicited in each area by low and high spatial frequency scenes. Also, a multivariate analysis was performed to test whether there are differences in category-specific information provided to each ROI by scenes containing only high or low spatial frequencies.

The results from the univariate analysis exhibited a very consistent trend across all of the category selective areas tested. While some of the comparisons were only marginally significant, every area responded more to HSF scenes than LSF scenes. PPA
was marginally more responsive to HSF scenes, but posterior PPA as well right PPA responded significantly more to HSF scenes, consistent with Rajimehr et al. (2011), but contrary to the findings of Peyrin et al. (2004, 2005).

Similar trends were observed in the multivariate analysis. It was not possible to decode the category of LSF scenes from any of the five main ROIs, but the HSF scenes were decoded above chance in all ROIs except for FFA. This result strongly suggests that category-specific information is lost when high spatial frequencies are removed from scene images. Interestingly, LSF scene category was decoded above chance from posterior PPA (but not anterior PPA) and there was a significant 2 way interaction with spatial frequency and PPA sub-region, suggesting that PPA might not be a functionally homogenous area as previously thought. Further work is needed in order to characterize the specific roles of anterior and posterior PPA in scene processing.

This study accomplished two main things that are worth emphasizing. First, it provides information regarding the spatial frequency tuning properties of several areas that have scarcely or never been examined in this manner. Second, it provides convincing evidence that high-level scene selective areas are primarily processing the finer structural details of scenes, portrayed by high spatial frequencies. At this stage of scene recognition, low spatial frequencies are not as important as previously theorized. This in no way suggests that low spatial frequency information is not essential in scene perception and categorization, only that high-level areas are not processing or using the coarse structure of the image by the time visual information reaches these areas of cortex.
There are numerous ways that subsequent research can use the information presented here and build on these studies. As far as the study detailed in chapter 1 is concerned, as retinotopic mapping techniques undergo minor adjustments or are completely overhauled, hopefully the information provided here can offer some useful guidelines in planning and executing more advanced methods. As far as the study detailed in chapter 2 is concerned, a variety of additional tests and analyses could be implemented in future studies. Obvious manipulations include changing the stimulus presentation times, filtering images at different spatial frequencies than the ones used here, and test different stimuli completely (such as faces or objects). Through these pursuits we can significantly advance our current understanding of the human visual system and resolve many of the unanswered questions regarding its underlying processes.
References


Dobre, C. M., Liu, G. T., Miki, A., & Modestino, E. J. (2001). Reproducibility of Retinotopic Mapping of Primary Visual Cortex Reproducibility of Retinotopic Mapping of Primary Visual Cortex. Children’s Hospital of Philadelphia Functional MRI Research Unit and the University of Pennsylvania School of Medicine,


96


98


99


Appendix A: Chapter 1 Supplementary information
A.1.

Fig A.1 The canonical hemodynamic response function \( \left( \frac{t}{p \cdot q} \right)^p \exp(p - \frac{t}{q}) \). The default values of \( p=8.6 \) and \( q=0.547 \) were used in this AFNI supplied HRF (Cohen, 1997). For the simulation, \( t \) (time in seconds) was designated as 16 seconds. For the fMRI experiment analysis the function was convolved with a square wave the same length as a single block and in the simulations this function was convolved with the design matrix (Fig. 3.1) which essentially replicates the GLM analysis performed using AFNI.
A.2. *Signal to noise ratio calculation*

The multifocal and phase-encoded retinotopic mapping methods analyze very different properties of neural activity and therefore require different computational treatment of the acquired data. Multifocal retinotopy uses a GLM approach to estimate the ordinary least squares (OLS) that minimizes the residual sum of squares (RSS) between predicted and observed neural responses. The GLM finds beta weights (the effects of interest in the predicted model) which statistical tests of significance (t-tests) and goodness of fit tests ($R^2$) can be performed on. On the other hand, phase-encoded retinotopic mapping requires a Fourier transform or cross-correlation analysis in order to find the harmonic most correlated with the time-series of responses measured at each voxel (Engel et al., 1997). Several statistical measures are generated from these computations including phase delays and signal coherence. Signal coherence is essentially a goodness of fit statistic for comparing different frequencies and is equivalent to squaring the cross-correlation coefficients calculated in time-series analyses (Yeragani et al., 2003).

Because the methodologies being examined here have these statistical incongruities, quantifying the efficiency of these experiments in a meaningful and directly comparable manner is a challenging proposition. Another issue is the fact that the ground truth cannot actually be determined, meaning that if the multifocal and phase-encoded retinotopic mapping experiments yielded different results, determining which method was more accurate might not be feasible. In order to similarly quantify the efficiency of both methods, an SNR statistic was calculated that denotes the extent in which the local
cortical representations corresponding to an individual region of visual field can be determined.

Contrary to the majority of retinotopic mapping analytical techniques that require working on the segmented cortical surface and smoothing the data, all of the computations of SNR calculations detailed here are performed in the original 3D space and no application of smoothing kernels. After a GLM analysis of the multifocal data was performed, marginal t-statistics and $R^2$ values were calculated on the beta coefficients of each voxel within an ROI. All marginal t-statistics were normalized to the peak activation for each regressor, after which a cluster analysis was performed. This was done by a region growing algorithm around the peak voxel location in the ROI. For each regressor these were considered the cluster of interest, or the “in” cluster. Next, the $R^2$ values within each cluster of interest were averaged and divided by the averaged $R^2$ values of all voxels in every other cluster. From this, a final weighted $R^2$ is reached and represents the SNR. A different procedure was used for estimating receptive fields in the phase-encoded data, where the phase estimates were binned according to their predicted visual field region so that the wedge and ring delay information could be intersected and the local region SNRs could be examined. Here, the squared cross-correlation coefficients within the each bin of interest were averaged and then divided by the average of all other squared cross-correlation coefficients of the other intersecting bins.

In the simulation, SNRs were obtained similarly for both the phase-encoded and multifocal retinotopic mapping methods, except that each voxel had a one-to-one
correspondence to a favored visual field position so that no region growing or binning of the data was necessary.

A.3. Receptive field estimation

The receptive fields pictured in figure 4.5a/b were fitted to individual voxels using the 2D Gaussian equation

\[ G(x, y) = \alpha \cdot \exp \left[ - \frac{1}{2(1 - \rho^2)} \left( \frac{(x - x_o)^2}{\sigma_x^2} + \frac{(y - y_o)^2}{\sigma_y^2} - \frac{2\rho(x - x_o)(y - y_o)}{\sigma_x \sigma_y} \right) \right] \]

Just as in the pRF method, the parameters \((x_0, y_0)\) describe Cartesian coordinates in which the receptive field is centered and the size is described by the standard deviation, \(\sigma\). For situations in which a local minimum could not be found during the fitting procedure a receptive field was not mapped. Constraints similar to the ones used in the pRF method were implemented, where the receptive field centers were contained within the radius of the stimulus and also to the corresponding hemifield.
Appendix B: Chapter 2 Supplementary information
For illustrative purposes, the canonical hemodynamic response function used in the spatial frequency fMRI experiment analyses defined as:

$$HRF(t) = \int_0^{\min(t,d)} s^q (g(t - s)) ds$$ where $$g(t) = \frac{t^q * \exp(-t)}{(q^q * \exp(-q))}$$

The default value of q=4 was used in this AFNI supplied HRF, ‘BLOCK’ (Cohen, 1997).
Supplementary Tables

Table B.1

Decoding Accuracies - Scene Category (4 Scenes)

<table>
<thead>
<tr>
<th>ROI</th>
<th>FS</th>
<th>LSF</th>
<th>HSF</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPA</td>
<td>39.38%*</td>
<td>28.13%</td>
<td>34.69%*</td>
</tr>
<tr>
<td>pPPA</td>
<td>33.44%*</td>
<td>30.31%*</td>
<td>33.75%*</td>
</tr>
<tr>
<td>aPPA</td>
<td>39.38%*</td>
<td>25.31%</td>
<td>40.00%*</td>
</tr>
<tr>
<td>RSC</td>
<td>31.25%</td>
<td>23.44%</td>
<td>33.44%*</td>
</tr>
<tr>
<td>OPA</td>
<td>33.75%*</td>
<td>23.13%</td>
<td>38.75%*</td>
</tr>
<tr>
<td>LOC</td>
<td>34.38%*</td>
<td>27.19%</td>
<td>36.25%*</td>
</tr>
<tr>
<td>FFA</td>
<td>30.94%*</td>
<td>25.63%</td>
<td>31.88%</td>
</tr>
<tr>
<td>Left PPA</td>
<td>34.69%*</td>
<td>30.00%*</td>
<td>37.19%*</td>
</tr>
<tr>
<td>Right PPA</td>
<td>34.69%*</td>
<td>26.25%</td>
<td>37.81%*</td>
</tr>
</tbody>
</table>

* Significant decoding accuracy above chance; p < 0.05.
Table B.2

Decoding of Spatial Frequencies -
(Binary) Scenes and Unstructured
Stimuli

ROI

<table>
<thead>
<tr>
<th></th>
<th>Scenes - Low/High</th>
<th>Unstructured - Low/High</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPA</td>
<td>66.72%*</td>
<td>55.94%</td>
</tr>
<tr>
<td>pPPA</td>
<td>64.69%*</td>
<td>60.63%*</td>
</tr>
<tr>
<td>aPPA</td>
<td>59.38%*</td>
<td>52.50%</td>
</tr>
<tr>
<td>RSC</td>
<td>52.19%</td>
<td>62.19%*</td>
</tr>
<tr>
<td>OPA</td>
<td>64.06%*</td>
<td>55.63%</td>
</tr>
<tr>
<td>LOC</td>
<td>62.81%*</td>
<td>57.81%</td>
</tr>
<tr>
<td>FFA</td>
<td>56.88%*</td>
<td>55.63%</td>
</tr>
<tr>
<td>Left PPA</td>
<td>64.06%*</td>
<td>51.25%</td>
</tr>
<tr>
<td>Right PPA</td>
<td>60.94%*</td>
<td>55.63%</td>
</tr>
</tbody>
</table>

* Significant decoding accuracy above chance; \( p < 0.05 \).