Decoding simulated neurodynamics predicts the perceptual consequences of age-related macular degeneration

Jianing V. Shi
Department of Biomedical Engineering, Columbia University, New York, NY, USA

Jim Wielaard
Department of Ophthalmology, Columbia University, New York, NY, USA

R. Theodore Smith
Department of Biomedical Engineering, Columbia University, New York, NY, USA

Paul Sajda
Department of Biomedical Engineering, Columbia University, New York, NY, USA

Age-related macular degeneration (AMD) is the major cause of blindness in the developed world. Though substantial work has been done to characterize the disease, it is difficult to predict how the state of an individual’s retina will ultimately affect their high-level perceptual function. In this paper, we describe an approach that couples retinal imaging with computational neural modeling of early visual processing to generate quantitative predictions of an individual’s visual perception. Using a patient population with mild to moderate AMD, we show that we are able to accurately predict subject-specific psychometric performance by decoding simulated neurodynamics that are a function of scotomas derived from an individual’s fundus image. On the population level, we find that our approach maps the disease on the retina to a representation that is a substantially better predictor of high-level perceptual performance than traditional clinical metrics such as drusen density and coverage. In summary, our work identifies possible new metrics for evaluating the efficacy of treatments for AMD at the level of the expected changes in high-level visual perception and, in general, typifies how computational neural models can be used as a framework to characterize the perceptual consequences of early visual pathologies.

Keywords: computational modeling, low vision, visual cortex


Introduction

Macular diseases such as age-related macular degeneration (AMD), diabetic retinopathy (DR), and macular dystrophy (MD) account for the overwhelming majority of blindness in the United States (Klaver, Wolfs, Vingerling, Hofman, & de Jong, 1998; Klein, Klein, & Linton, 1992). Approximately 15 million people in the United States have some signs of macular degeneration, with the rate of new cases dramatically increasing due to longer life expectancies and the aging baby-boomer population. Macular disease is not limited to older generations. It is also a common problem for diabetics in the United States, 7 million of whom suffer from diabetic retinopathy.

Current efforts for tracking and treating macular disease have focused on the retina, for instance, quantification of drusen distributions, photodynamic therapy (Wormald, Evans, Smeeth, & Henshaw, 2003), and even retinal prostheses for degenerations of the entire retina (Brindlay & Lewin, 1969; Humayun et al., 2003; Zrenner, 2002). Color photographic images are commonly used for the diagnosis, treatment, and staging of AMD and other macular diseases. Early lesions imaged by fundus photography are the subretinal deposits known as drusen and abnormalities of the retinal pigment epithelium (RPE; Bressler, Bressler, Seddon, Gragoudas, & Jacobson, 1998; Bressler, Maguire, Bressler, & Fine, 1990; Smiddy & Fine, 1984). The late lesions, usually accompanied by severe vision loss, are geographic atrophy (GA) and choroidal neovascularization (CNV; Bressler, Bressler, & Fine, 1988; Sunness, Bressler, Tian, Alexander, & Applegate, 1999; Sunness, Gonzalez-Baron, Bressler, Hawkins, & Applegate, 1999). Besides photography, the eye permits other optical imaging technologies using the scanning laser ophthalmoscope (SLO; Kirkpatrick,
The SLO can image fundus autofluorescence, the source of which is the lipofuscin in the RPE, and is an important marker in retinal degenerations (Delori, Fleckner, Goger, Weiner, & Dorey, 2000; Smith et al., 2006; von Ruckmann, Fitzke, & Bird, 1997). The SLO can also acquire infrared reflectance scans (Beausencourt, Remky, Elsner, Hartnett, & Trempe, 2000; Elsner, Burns, Weiner, & Delori, 1996), which can reveal other RPE defects, such as drusen and edema. Because tissue sampling of living retinas is quite hazardous, the information in these images is of paramount importance for clinical assessment.

Retinal imaging, however, does not provide a complete picture for the nature of the expected vision loss. It is important to consider how the visual cortex responds to the distortion of the retinal input and how this relates to perception. Psychophysics has played an important role in characterizing the effects of retinal scotomas on visual perception. Early efforts focused on perceptual processes for filling-in a scene across the retinal blind spot. For instance, Ramachadrara (1992) showed perceptual filling-in of background color, bars, and geometric patterns at the blind spot. Other studies have shown that the filling-in of the blind spot in one eye can influence perception arising from the other eye (Murakami, 1995; Tripathy & Levi, 1994). This filling-in process has also been shown to take place early in perceptual processing (Murakami, 1995) and induces little or no distortion of the surrounding region at the blind spot (Tripathy, Levi, & Ogmen, 1996). Studies by Kawabata (1982, 1984) have shown that complex patterns such as gratings, concentric circles, and dotted lines can be filled in across the blind spot. A recent study using a computational model has shown a possible mechanism for perceptual filling-in following retinal degeneration (McManus, Ullman, & Gilbert, 2008).

We take a different approach from these efforts that explicitly investigate the role of filling-in phenomena in AMD. Instead, we directly consider the question of how the visual cortex responds to the distortion of the retinal input and how such changes in neural firing might affect decoding and ultimately high-level perception. Specifically, we use a large-scale anatomically and physiologically realistic spiking neuron model of LGN and layer 4 in primary visual cortex (V1) as a substrate for mapping the retinal input into network neurodynamics. This model has been described previously (Wiegaard & Sajda, 2006a, 2006b) and has been shown to generate a substantial fraction of the classical and extraclassical response properties seen experimentally at both the single cell and population level. A hallmark of the model is that there are no long-range intra-cortical connections. Thus, our focus is to investigate how retinal pathologies, such as macular drusen, affect the representation of the stimulus in terms of the early input layers of V1.

In order to map the neurodynamics to simulated perception, we utilize an approach from signal detection theory that has been extensively used in systems neuroscience to study perceptual decision making (e.g., see Britten, Shadlen, Newsome, & Movshon, 1992; Philiaistides & Sajda, 2006). Specifically, we construct neurometric functions by decoding neurodynamics that have been perturbed by particular pathologies seen in an individual’s fundus image. These neurometric functions are then compared to the specific individual's psychometric performance. We analyze the quality of these predictions relative to more standard metrics based on the statistics of the drusen on the retina.

**Methods**

**Patient recruitment and psychophysics experiment**

We recruited 10 low-vision patients with mild yet progressive macular degeneration, as well as 10 age-matched healthy controls at the Edward Harkness Eye Institute, Columbia Presbyterian Medical Center. All participants provided written informed consent, as approved by the Columbia University Institutional Review Board. All subjects, whose ages ranged from 65 to 84, had 20/20 to 20/50 corrected visual acuity. The subjects we term as “controls” had healthy vision in at least one of their eyes, with corrected visual acuity of 20/20. All psychophysics tests were conducted monocularly.

We used a two-alternative forced-choice (2-AFC) face versus car visual discrimination paradigm. Impaired face perception is one of the disabilities caused by AMD (Bullimore, Bailey, & Wacker, 1991) with those affected reporting it as one of their most significant complaints (McClure, Hart, Jackson, Stevenson, & Chakravarthy, 2000). We used a set of 12 faces from the Max Plank Institute face database (Troje & Bulthoff, 1996) and 12 car grayscale images. The car image database was the same used in Philiaistides and Sajda (2006) and Philiaistides, Ratcliff, and Sajda (2006) and, in summary, was constructed by taking images from the Internet, segmenting the car from the background, converting the image to grayscale, and then resizing to be comparable in size to the face images. The pose of the faces and cars was also matched across the entire database and was sampled at random (left, right, center) for the training and test cases. Varying the pose of the objects in the images was to ensure that both human subjects and the model exhibited some pose invariance in their decoding and thus were less likely to be classified based on accidental features—e.g., accidental features due to the direction of illumination.

All images were 512 x 512 pixels, 8 bits/pixel, and were equated for spatial frequency, luminance, and contrast. The phase spectra of the images were manipulated using the weighted mean phase method (Dakin, 2002; Philiaistides & Sajda, 2006) to modulate the decision difficulty, resulting
in a set of images graded by phase coherence. Sample images are shown in Figure 1.

The sequence of images was presented to subjects in a random design, where each image was flashed for 50 ms and images subtended 2° × 2° of visual angle with the screen background set to the mean luminance gray. Subjects were instructed to fixate on a fixation cross between trials. To set the interstimulus interval (ISI), we ran a set of pilot experiments to determine a reasonable ISI for these sets of subjects. Our prior psychophysics work for this paradigm, in which we had a younger population of subjects having normal to corrected-to-normal vision (see Philiaistides & Sajda, 2006; Philiastides et al., 2006), found that ISIs of 1500–2000 ms were of sufficient duration to enable a rapid response without having subjects generate substantial misses due to a lack of time to respond. Since our subject population is substantially older with poorer vision, it is not surprising that our pilot psychophysics experiments showed that the 1500–2000 ms ISI was too short for this population, as seen by a large fraction of missed stimuli (i.e., subjects did not make a response prior to the next stimulus). We extended the ISI range by 1 s (in the range of 2500–3000 ms) and found that miss rates were substantially reduced and more in line with the younger, normal vision subject population. Subjects performed 24 trials per coherence level, with coherence levels of 20%, 25%, 30%, 35%, 40%, 45%, and 55%. A Dell computer with NVIDIA GeForce4 MX 440 with AGP8X graphics card and E-Prime software controlled the stimulus presentation.

We used a similar experimental paradigm to conduct the simulated experiments with the network model. The sequence of images was also presented to the model (see below for the description of the model) in a random design, where an image was flashed for the same duration of 50 ms. However, different from the human psychophysics experiments, we used an ISI of 200 ms for our simulations. Since simulating the model is computationally expensive, we minimized the simulation time by choosing an ISI that was as small as possible yet did not result in network dynamics leaking across trials. We conducted pilot experiments which showed that network activity settled to background levels approximately 200 ms after stimulus offset (i.e., after the stimulus was removed from the input), which led to our choice of 200 ms for the simulation ISI. We ran the simulations with each trial randomized for image class

Figure 1. Summary of the perceptual decision-making experimental design, a two-alternative forced-choice paradigm for face versus car discrimination. Images were flashed for 50 ms, followed by an interval of 200 ms with the same mean luminance as the stimulus. By varying the phase coherence, we manipulated the evidence in the stimuli for face or car. We used the same set of stimuli for the human psychophysics experiment and V1 model simulation. Examples of face and car images at each of the coherences used in the experiments are shown.
(face or car) and coherence level, respectively—i.e., same as for human psychophysics. Each image (class and coherence) was repeated 30 times in each of the simulations.

Fundus image analysis

After pupillary dilation, color fundus and red-free (RF) fundus photographs were either taken using film-based photographs acquired on with a Topcon TRC-50EX (Topcon Medical Systems, Paramus, NJ, USA) and digitized on the Nikon CoolScan V (Nikon, Tokyo, Japan) or acquired digitally on the Zeiss FF 450 Plus Fundus Camera (Carl Zeiss Meditec, Jena, Germany).

Segmentation of drusen was performed on the RF fundus photographs using a robust and automated algorithm (Smith et al., 2005), with sample results illustrated in Figure 2. A background was constructed for each image, guided by the specific anatomy of the principal absorbers and reflectors. After background subtraction, an automated algorithm was used to segment the drusen in the fundus images.

Modeling retinal impairment

Among the patient eye images we collected, we excluded the wet macular degeneration cases from our analysis. Wet macular degeneration is a more severe form of AMD than the dry form and accounts for approximately 10% of all AMD but 90% of all blindness from the disease. The wet form is characterized by choroidal neovascularization (CNV), the development of abnormal blood vessels beneath the retinal pigment epithelium (RPE) layer of the retina. These vessels can bleed and cause macular scarring that further result in profound loss of central vision, which would be beyond the scope of our retinal model. We found after testing that one eye among the ten patient eyes had the wet form of AMD.

Among the nine dry AMD cases, both hard drusen and soft drusen were observed. The hard drusen are characteristic of earlier stages of macular degeneration and appear as yellow, discrete spots on color fundus photographs between 1 and 63 µm. The soft drusen, on the other hand, are found in earlier and late-stage macular degeneration, typically associated with pigmentary changes as disease progresses. The soft drusen exhibit yellow, fuzzy appearance on color fundus photographs, which are above 125 µm or from 63 to 125 µm with visible thickness (Klein et al., 1991).

We modeled the retinal impairment given the drusen segmentation for the dry AMD cases. A simple thresholding operation was used to construct the binary mask:

$$\rho^\text{AMD}(\vec{y}) = \begin{cases} 0 & \text{if } \vec{y} \in \text{drusen areas}, \\ 1 & \text{if } \vec{y} \notin \text{drusen areas}, \end{cases}$$

where $\rho^\text{AMD}(\vec{y})$ denotes the binary mask and $\vec{y}$ is the spatial location on the binary mask. As an approximation, we treated all the drusen as scotomas and constructed binary masks by thresholding the segmented fundus images. Such an assumption tests the limit of our cortical model and generates a first-order prediction for the perceptual vision loss. We also assumed that all the patients used their fovea as their preferred retinal location. Figure 3 illustrates the approach, where drusen acts as a binary mask on the retinal input.

Simulating cortical activities using a model of V1

We used an anatomically and physiologically realistic model of V1 model to simulate the cortical consequences of retinal impairment. Details of the V1 model have been
A sparse linear decoder maps the population spike trains into a decision.

The impaired visual input is fed into the large-scale model of V1 and a sparse linear decoder to map the retinal impairment into cortical and perceptual space. The binary mask was used to modulate the input conductance from LGN to V1 by acting multiplicatively on the visual stimulus. In the example shown above, a face image extends over 2° × 2° of visual field and passes through the binary mask bounded by the red square. The impaired visual input is fed into the large-scale model of V1. A sparse linear decoder maps the population spike trains into a decision.

described previously (McLaughlin, Shapley, Shelley, & Wielaard, 2000; Wielaard & Sajda, 2006a, 2006b). In brief, the model consists of a layer of N (4096) conductance-based integrate-and-fire point neurons (one compartment), representing about a 2° × 2° piece of a V1 input layer (layer 4C). Our model of V1 consists of 75% excitatory neurons and 25% inhibitory neurons. Dynamic variables of each neuron are the membrane potential $v_i(t)$ and spike train $S_i(t) = \sum \delta(t - t_{ik})$, where $t$ is time and $t_{ik}$ is the $k$th spike of the $i$th neuron, $i = 1, \ldots, N$. Each neuron is modeled as

$$C_i \frac{dv_i}{dt} = -gL_i(v_i - V_L) - gE_i(v_i - V_E) - gI_i(v_i - V_L),$$

(2)

where the quantities $gL_i$, $gE_i$, and $gI_i$ represent the leakage, excitatory, and inhibitory conductances of neuron $i$.

Both the excitatory and inhibitory populations consist of two subpopulations $P_k(E)$ and $P_k(I)$, $k = 0, 1$, a population that receives LGN input ($k = 1$) and one that does not ($k = 0$). In the model, 30% of both the excitatory and inhibitory cell populations receive LGN input. Noise, cortical interactions, and LGN input are assumed to act additively in contributing to the total conductance of a cell:

$$g_{E,i}(t) = \left\{ \begin{array}{ll}
\eta_{E,i}(t) + g_{E,i}^{\text{cor}}(t[S]_E) & \text{if } i \in \{P_0(E), I\} \\
\eta_{E,i}(t) + g_{E,i}^{\text{cor}}(t[S]_E) + g_{E,i}^{\text{LGN}}(t) & \text{if } i \in \{P_1(E), I\},
\end{array} \right.$$  

$$g_{I,i} = \eta_{I,i}(t) + g_{I,i}^{\text{cor}}(t[S]_I) \forall i \in \{1, \ldots, N\},$$

(3)

where $\eta_{\mu,i}(t)$ represents a cell-specific external stochastic term representing synaptic noise for a cortical excitatory ($\mu = \text{E}$) or inhibitory ($\mu = \text{I}$) neuron (see Wielaard & Sajda, 2006b and Supplementary materials for details). The terms $g_{\mu,i}^{\text{cor}}(t[S]_\mu)$ are the contributions from the cortical excitatory ($\mu = \text{E}$) and inhibitory ($\mu = \text{I}$) neurons and include only isotropic connections:

$$g_{\mu,i}^{\text{cor}}(t[S]_\mu) = \int_{-\infty}^{+\infty} ds \sum_{k=0}^{1} \sum_{i \in \{P_1(\mu)\}} C_{i,k}^{\text{cor}}(||\vec{x}_i - \vec{x}_j||) \cdot G_{\mu,j}(t - s)S_j(s),$$

(4)

where $i \in \{P_1(\mu)\}$. Here, $\vec{x}_i$ is the spatial position (in cortex) of neuron $i$, the functions $G_{\mu,j}(\tau)$ describe the synaptic dynamics of cortical synapses, and the functions $C_{i,k}^{\text{cor}}(r)$ describe the cortical spatial couplings (cortical connections). The length scale of excitatory and inhibitory connections is about 200 μm and 100 μm, respectively.

In the model, 30% of all neurons receive LGN input. In agreement with experimental findings, the LGN neurons are modeled as rectified center–surround linear spatiotemporal filters. A cortical cell, $j \in \{P_1(\mu)\}$, is connected to a set $N_{\text{LGN}}^L$ of left eye LGN cells or to a set $N_{\text{LGN}}^R$ of right eye LGN cells:

$$g_{LGN}^j(t) = \sum_{e \in N_{\text{LGN}}^L} \left[ g_0^e + g_1^e \int_{-\infty}^{+\infty} ds \int d^2 \vec{y} G_{LGN}^e(t - s) \cdot \mathcal{L}_e(||\vec{y}_i - \vec{y}_j||) \rho_{\text{AMD}}(\vec{y}) I(\vec{y}, s) \right],$$

(5)

where $Q = \text{L}$ or $\text{R}$ (i.e., left or right eye). Here, $[x]_+ = x$ if $x \geq 0$ and $[x]_+ = 0$ if $x \leq 0$, $\mathcal{L}_e(r)$ and $G_{LGN}^e(t)$ are the spatial and temporal LGN kernels, respectively, $\vec{y}_i$ is the receptive field center of the $i$th left or right eye LGN cell, which is connected to the $j$th cortical cell, and $I(\vec{y}, s)$ is the visual stimulus. The parameters $g_0^e$ represent the maintained activity of LGN cells and the parameters $g_1^e$ measure their responsiveness to visual stimuli. The binary mask $\rho_{\text{AMD}}(\vec{y})$ occludes the visual stimulus $I(\vec{y}, s)$ at the location of scotoma, which acts multiplicatively on the input conductance to the V1 model.
The LGN kernels are of the form:

\[
G_{\ell}(t) = \begin{cases} 
0 & \text{if } t \leq t_0^\ell \\
K_{\ell} \tau_t e^{-\tau_t/(\tau_1 t_1) - c e^{-\tau_t/(\tau_2 t_2)}} & \text{if } t > t_0^\ell
\end{cases}
\]

and

\[
\mathcal{L}_{\ell}(r) = \frac{z(1 - K_{\ell})^{-1}}{2} \left\{ \frac{1}{\pi \sigma_z e^{-(r/\sigma_z)^2}} - \frac{K_{\ell}}{\pi \sigma_c e^{-(r/\sigma_c)^2}} \right\},
\]

where \( k \) is a normalization constant, \( \sigma_c, \ell \) and \( \sigma_z, \ell \) are the center and surround sizes, respectively, and \( K_{\ell} \) is the integrated surround–center sensitivity. The temporal kernels are normalized in Fourier space, \( \frac{1}{2\pi} \int_{-\infty}^{\infty} |G_{\ell}(\omega)| d\omega = 1 \). 

\[ G_{\ell}(\omega) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} G_{\ell}(t) e^{-i\omega t} dt. \]

For the magnocellular architecture, the time constants \( \tau_1 = 2.5 \text{ ms}, \tau_2 = 7.5 \text{ ms}, \text{ and } c = (\tau_1/\tau_2)^6 \) so that \( G_{\ell}(0) = 0 \), in agreement with the experiments (Benardete & Kaplan, 1999). For the parvocellular architecture, the time constants \( \tau_1 = 8 \text{ ms}, \tau_2 = 9 \text{ ms}, \text{ and } c = 0.7(\tau_1/\tau_2)^5 \). The delay times \( t_0^\ell \) are taken from a uniform distribution between 20 ms and 30 ms, for all cases. Sizes for center and surround were taken from experimental data (Croner & Kaplan, 1995; Derrington & Lennie, 1984; Hicks, Lee, & Vidyasagar, 1983; Shapley, 1990; Spear, Moore, Kim, Xue, & Tumosa, 1984) and were \( \sigma_c, \ell = \sigma_c = 0.1^\circ \) (magn) and 0.04° (parv) for centers and \( \sigma_z, \ell = \sigma_z = 0.72^\circ \) (magn) and 0.32° (parv) for surrounds. The integrated surround–center sensitivity was in all cases \( K_{\ell} = 0.55 \) (Croner & Kaplan, 1995). By design, no diversity has been introduced in the center and surround sizes in order to demonstrate the level of diversity resulting purely from the cortical interactions and the connection specificity between LGN cells and cortical cells (i.e., the sets \( N_{Q,J} \), see specifications below). Furthermore, no distinction was made between ON-center and OFF-center LGN cells other than the sign reversal of their receptive fields (± sign in Equation 7). The LGN RF centers \( \bar{y}_t \) were organized on a square lattice with lattice constants \( \sigma_c/2 \). These lattice spacings and consequent LGN receptive field densities imply LGN cellular magnification factors that are in the range of the experimental data available for macaque (Conolly & van Essen, 1984; Malpeli, Lee, & Baker, 1996). The connection structure between LGN cells and cortical cells, given by the sets \( N_{Q,J} \), is made so as to establish ocular dominance bands and a slight orientation preference that is organized in pinwheels (Blasdel, 1992).

It is further constructed under the constraint that the LGN axonal arbor sizes in V1 do not exceed the anatomically established values of 1.2 mm for magnocellular and 0.6 mm for parvocellular neurons (Blasdel & Lund, 1983; Freund, Martin, Soltesz, Somogyi, & Whitteridge, 1989).

In the construction of the model, our objective was to keep the parameters deterministic and uniform as much as possible. This enhances the transparency of the model while at the same time provides insight into what factors may be essential for the considerable diversity observed in the responses of V1 cells. Important parameters that are not subject to cell-specific variability are:

1. Parameters related to the integrate-and-fire mechanism, such as threshold, reset voltage, and leakage conductance. These are identical for all cells (Equation 2).
2. The cortical interaction strengths and connectivity length scales. These are presented by the functions \( C_{\mu,k}(r) \) that are not cell specific but only specific with respect to the four cell populations. Note that the functions \( C_{\mu,k}(r) \) are also not configuration specific (Equation 4).
3. Maintained activity and responsiveness to visual stimulation of LGN cells (Equation 5).
4. Receptive field sizes of LGN cells. These are neither cell nor population specific (i.e., where “population” in this case refers to the ON and OFF LGN cell populations) but are only specific with respect to the four model configurations, i.e., receptive field sizes of all LGN cells are identical for a particular configuration (Equation 7).

Important parameters that are subject to cell-specific variability are:

1. The external noisy conductances \( \eta_c,\ell(t) \) (excitatory) and \( \eta_s,\ell(t) \) (inhibitory) (Equation 3).
2. The cortical synaptic dynamics as described by the kernels \( \xi_{\mu,\ell}(t) \) (Equation 4).
3. The LGN temporal kernels \( G_{\ell}(r) \) (Equation 5).
4. The LGN connectivity to our model cortex as described by \( N_{L,J} \) and \( N_{K,J} \) (Equation 5).

Additional details of model parameters and tuning can be found in Wielaard and Sajda (2006b).

In order to characterize the tuning properties of the cortical neurons under retinal impairment, we simulated the cortical activities using drifting grating stimuli, for both magnocellular and parvocellular versions of the model. The model used in simulating the orientation tuning experiment has 128 × 128 neurons. We characterized the sharpness of orientation tuning for cortical neurons on a population basis.

For the 2-AFC simulations for the model, we decoded cortical activities for the face versus car discrimination task using a medium-sized cortical model to accelerate simulation. The model used in the discrimination task had 64 × 64 neurons (4096 total neurons). We used this reduced sized model, rather than the large 128 × 128 model, since...
it reduced simulation time by several orders of magnitude. Pilot experiments also showed that the classical and extra-classical responses generated using the smaller model were not significantly different from the larger model.

**Decoding neural activity**

We used a linear decoder to map the spatiotemporal activity in the recurrent V1 model to a binary decision (e.g., face or car). For the purpose of this work, we use the term “linear decoder” to refer to a two-class discriminative linear classifier, in which the decoder learns a linear hyperplane in the input/feature space that maps inputs into one of two classes (e.g., face or car). Our justification for linearly decoding neural activity from the early visual pathways is based on both theoretical and experimental grounds. One hypothesis on how the ventral stream is able to implement invariant object recognition is through “manifold untangling” (DiCarlo & Cox, 2007). Conceptually, the hypothesis asserts that the visual stimulus is mapped, via the ventral stream, into a space of neurodynamics in which manifolds that represent different object class are “flattened out.” In the manifold space, object classes can then be separated via (linear) hyperplanes. A similar linear decoding approach has recently been applied to decoding neural activity of neuronal populations in macaque primary visual cortex (Graf, Kohn, Jazayeri, & Movshon, 2011), though in that case firing rates over long time periods in the trial were used, whereas here (see below) we consider firing rates computing over short and long time windows (i.e., spike counts for short and long time bins). We imposed a sparsity constraint on the linear decoder to control the dimension of the feature space (which is the spike count matrix—see below). Sparse decoding has been applied to a variety of neurophysiological data (Chen, Geisler, & Seidemann, 2006; Palmer, Cheng, & Seidemann, 2007; Quiroga, Reddy, Koch, & Fried, 2007; Quiroga, Reddy, Kreiman, Koch, & Fried, 2005).

Specially, our linear decoding begins by constructing the spike train for each neuron $i$, in the population, for each trial $k$, as $s_{i,k}(t) = \sum_\delta(t - t_{i,j,k})$, where $t \in [0, 250]$ ms, $i = 1 \ldots N$ is the index for neurons, $k = 1 \ldots M$ is the index for trials, and $l = 1 \ldots P$ is the index for spikes. Based on the population spike trains, we estimated the firing rate on each trial by counting the number of spikes within a time bin of width $\tau$, resulting in a “spike count matrix” $r_{i,j,k} = \int_{(j-1)\tau}^{j\tau} s_{i,k}(t)dt$, where $i = 1 \ldots N$ is the index for neurons, $j = 1 \ldots T/\tau$ is the index for time bin, and $k = 1 \ldots M$ is the index for trials. When $\tau = 25$ ms, we are assuming that information is encoded in the temporal precision of the population activity since temporal precision is required so that the spike count matrix does not, from trial to trial, change substantially by having spikes switch from one bin to another. When $\tau = 250$ ms, we integrate the spiking activity over the entire trial, leading to a rate-based representation of information.

The class labels of each sample $b \in \mathbb{R}^m$ take the value of $\{-1, +1\}$ (either face or car). We then compute the weighted sum over the population spike count matrix. For notational convenience, we replace the spike count matrix $r_{i,j,k}$ with the stacked matrix $x_{i,k}$, where $x_{i,k} = r_{i,j,k}$, and $l = (i - 1)n + j$, which leads to the following constrained minimization problem:

$$\{w, v\} = \arg\min_M \frac{1}{M} \sum_{i=1}^M \theta((w^T x_i + v) b_i) + \lambda \|w\|_1, \quad (8)$$

Here, $\lambda > 0$ is a regularization parameter that controls the sparsity of the decoder, $w \in \mathbb{R}^d$ specifies the weights, and $v \in \mathbb{R}$ is the offset/bias. The parameter $\theta$ is the logistic loss function, defined by $\theta(z) = \log(1 + \exp(-z))$. Such a formulation essentially minimizes the average logistic loss defined by the first term in the minimization, with a Lagrange multiplier for the $\ell_1$ norm of the weights—i.e., reduces the classification error while choosing as few elements in the spike count matrix as possible. The resultant linear decoder can be geometrically interpreted as a hyperplane defined by $w^T X + v = 0$, which separates the classes of face and car. We optimize Equation 8 using the hybrid iterative shrinkage (HIS) algorithm (Shi, Yin, Osher, & Sajda, 2010).

Training and testing were carried out on different sets of images, each containing 6 face images and 6 car images, 30 trials per image. We performed training and testing on each individual phase coherence independently. $K$-fold cross-validation (where $K = 10$) was used on the training set, while the weights applied on the testing set were estimated using a jackknife estimation to eliminate the bias.

**Construction and statistical tests of neurometric and psychometric curves**

Psychometric curves were constructed by fitting a cumulative Weibull function (Quick, 1974) to the coherence vs. percent correct behavioral data. Neurometric curves were constructed by first estimating the area under the receiver operating characteristic (ROC) curve (area under this curve is $Az$) for the $K$-fold results of the model. $Az$ can be seen as the probability of a correct decision by the decoder (Green & Swets, 1966). These data were then also fit using a Weibull function.

We used a likelihood ratio test (Hoel, Port, & Stone, 1971) to quantify the degree of similarity between the psychometric and neurometric functions. We do this by fitting the best single Weibull function jointly to the two data
Figure 4. Simulating cortical activity and orientation tuning across the cortical network for baseline (no drusen) vs. AMD subjects (drusen constructed using individual fundus images). For (A), (B), (D), and (E), the 4096 simulated cortical neurons ($64 \times 64$ neurons), arranged as $8 \times 8$ orientation hypercolumns, are shown. The vertical banding is due to the fact that the stimulus is monocular (in both model simulations and human psychophysics) and thus highlights the ocular dominance column structure in the network. (Top row) Results for magnocellular cortex. (A, D) Firing rates for a drifting grating stimulus for a control (no drusen) simulation. (B, E) Average firing rates for a simulated AMD (drusen) case. (C, F) Distribution of orientation tuning, as measured via CV, across all neurons. The red curve denotes the CV distribution for the control and curves of other colors indicate the CV distribution for simulated AMD patients. (Bottom row) Same as top except results are for parvocellular architecture. Note that in both cases this represents monocular input simulations, and therefore, the vertical banding represents the ocular dominance columns in the model. For (A), (B), (D), and (E), each grid point represents a simulated cortical neuron, with the color indicating its firing rate as specified by the color bar legend.
sets in addition to the individual fits. The likelihoods \( L_s \) obtained from these two conditions were transformed by

\[
\lambda = -2 \ln \frac{L(\text{data jointed curve})}{L(\text{data individual curves})},
\]

so that \( \lambda \) is distributed as \( \chi^2 \) with 2 degrees of freedom (Hoel et al., 1971). If \( \lambda \) does not exceed the criterion value (for \( p = 0.05 \)), we conclude that we cannot reject the hypothesis that a single function fits the two data sets as well as two separate functions.

\[
CV = 1 - \left[ \frac{\int r(\theta) \exp(2i\theta) d\theta}{\int r(\theta) d\theta} \right],
\]

where \( r(\theta) \) is the mean firing rate at orientation \( \theta \in [0, 2\pi] \). CV is a measure of orientation selectivity, where a smaller value for CV indicates a greater orientation selectivity. When \( CV = 0 \), the neuron only responds to one orientation; while \( CV = 1 \), the neuron responds equally to all orientations and, hence, is not selective for orientation. Orientation selectivity is one of the fundamental properties of the early visual system and is a major element of form vision needed for object recognition and discrimination.

We investigated both magnocellular and parvocellular versions of the model (Wielaard & Sajda, 2006b) comparing them in terms of basic firing rate and tuning characteristics. Interestingly, these two architectures respond differently to impairment of retinal input defined by drusen patterns. Specifically, we see differences in both firing rates of cortical neurons and their orientation selectivity. Figure 4 (top row) illustrates the cortical responses in the magnocellular cortex. The firing rates are reduced overall for neurons in the magnocellular system, though the spatial distribution of active neurons remains largely unaffected. However, the orientation selectivity of the network is significantly affected, with the distribution of CV consistently shifted to the right, indicating that neurons in the magnocellular system become less orientation selective when the stimulus is masked by drusen.

Figure 5. Psychometric curves for both control subjects (red) and AMD patients (blue), constructed from behavioral data. Both curves represent average psychometric performance across subjects. Clearly, the AMD patients suffer from lower discrimination accuracy compared to control subjects. The degradation of patient performance is more pronounced at higher phase coherences. Error bars indicate standard error.

Figure 6. Comparing predicted neurometric curves with group average psychometric curve. The average psychometric curve for 9 AMD patients (blue: same curve as in Figure 2) is plotted, together with corresponding neurometric curves (black and gray), computed from drusen data from the same 9 AMD patients. The black neurometric curve is constructed using a temporally narrow binning of neuronal activity (25 ms bins), capturing temporally fine neurodynamics. The gray neurometric curve is constructed using a large temporal bin (250 ms) and thus represents a temporally integrated response more like a rate-based code for each trial. Note that, for all curves, the error bars indicate standard error.

**Results**

Cortical activity and tuning in the presence of retinal scotoma

We began by evaluating classical tuning properties of the model for baseline (no drusen) and AMD (drusen) conditions. Specifically, we characterized the orientation selectivity of modeled cortical neurons by using drifting sinusoidal gratings to measure the circular variance (CV; Ringach, Shapley, & Hawken, 2002), which is defined as
Figure 7. Three patient cases (A–C), where the neurometric curves are good predictors for the individual's psychometric curves, are illustrated. (Left) Red-free (RF) fundus image of the patient. (Middle) Binary retinal mask used as input to the V1 model. The red square indicates the area on the retinal mask that is fed into the V1 model. (Right) Comparison of the neurometric curve (thick black) and the corresponding psychometric curve of the given patient (thick blue) plotted against the individual psychometric curves (dashed blue) for all the patients. For these three subjects, psychophysical and neuronal data were statistically indistinguishable as assessed by a likelihood ratio test after we fit the best single Weibull function jointly to the two data sets. The $p$-value in each panel represents the results of this test. A $p$-value > 0.05 indicates that there is no significant difference between a fit to the data using two separate functions and that using a single function.
Figure 8. Three patient cases (A–C), where the neurometric curves are different from the psychometric curves, are illustrated. (Left) Red-free (RF) fundus image of the patient. (Middle) Binary retinal mask used as input to the V1 model. The red square indicates the area on the retinal mask that is fed into the V1 model. (Right) Comparison of the neurometric curve (thick black) and the corresponding psychometric curve of the given patient (thick blue) plotted against the individual psychometric curves (dashed blue) for all the patients. For two of the three subject cases (B, C), we could reject the null hypothesis, at $p < 0.05$, that a single curve predicts both the neurometric and psychometric functions. For subject A, the psychometric and neurometric curves are clearly different; however, differences are not significant at $p < 0.05$. 
Figure 4 (bottom row) shows the cortical responses of the parvocellular architecture. Unlike the magnocellular system, there are “holes” of activity—i.e., patches of simulated cortex in which activity is substantially reduced relative to no drusen mask. The spatial distribution of such inactivation is correlated with the location of drusen in the visual field. Orientation selectivity of the neurons, on the other hand, is not significantly affected. From these results, we can infer that the magnocellular system performs some amount of “filling-in” of cortical activity, without the need to have long-range cortical connections. Such filling-in processes, which are seen as one of the mechanisms the visual system uses to compensate for scotoma, particularly in AMD (McManus et al., 2008; Zur & Ullman, 2003), naturally arise from the model’s receptive field scatter and short-range connectivity.

Figure 4 (bottom row) shows the cortical responses of the parvocellular architecture. Unlike the magnocellular system, there are “holes” of activity—i.e., patches of simulated cortex in which activity is substantially reduced relative to no drusen mask. The spatial distribution of such inactivation is correlated with the location of drusen in the visual field. Orientation selectivity of the neurons, on the other hand, is not significantly affected. From these results, we can infer that the magnocellular system performs some amount of “filling-in” of cortical activity, without the need to have long-range cortical connections. Such filling-in processes, which are seen as one of the mechanisms the visual system uses to compensate for scotoma, particularly in AMD (McManus et al., 2008; Zur & Ullman, 2003), naturally arise from the model’s receptive field scatter and short-range connectivity.

It has been suggested that holistic face perception largely involves low spatial frequencies (LSF; Goffaux & Rossion, 2006). LSF, the filling-in results demonstrated above, and the transient nature of the stimulus presentation all point to the magnocellular pathway as being the most appropriate for mapping the stimulus to the model’s neurodynamics. Thus, all our subsequent analysis and results are reported using simulations from the magnocellular model.
Figure 10. Distribution of orientation tuning (measured via circular variance: CV) and simple and complex cell distributions (measured via the modulation ratio: F1/F0) for (A) full magnocellular cortical network used in simulations and (B) informative neurons selected for the drusen cases in Figure 7 (same ordering as in Figure 7).
loss in visual perception, is consistent with clinical data (Remky & Elsner, 2005).

Figure 8 shows another three cases where the neuro-
metric curves diverge from the psychometric curves. There are several factors that could explain why these predictions are not as accurate as those in Figure 7. In Figures 8A and 8B, the part of the visual field seen by the model (indicated by the red square) has a very different pattern of drusen than the rest of the fundus. Thus, the simulations may not capture the true retinal pathologies, particularly if the subjects’ had a preferred retinal locus (PLR) that was significantly off-fovea or even outside the $2^\circ \times 2^\circ$ visual field captured by the simulations. Figure 8C is a case of reticular macular disease (Smith, Sohrab, Busuioc, & Barile, 2009) in which the severity of the disease and vision loss is not well characterized simply by looking at drusen density and coverage.

We next analyzed several of the properties of the net-
work that affect decoding performance. We first con-
sidered the “informative neurons” utilized by the decoder to map neural activity to a perceptual decision. We define “informative neurons” as those neurons selected by the decoder given that they have at least one non-zero weight. Figure 9 shows the results for a sample AMD case and also a case with no retinal drusen. In both cases, the number of informative neurons increases as the signal-to-noise ratio (SNR) of the stimulus decreases and the decision becomes more difficult. This reflects the decoder recruiting more neurons as the SNR decreases. We also find that the difference between the AMD and no drusen cases is an approximate constant increase in informative neurons, across all coherence levels—i.e., an additional 100–200 neurons are recruited in the AMD case vs. the control case, regardless of the coherence level of the stimulus.

We next considered the tuning of the informative neurons relative to the tuning of all cells in the network model. Figure 10 shows distributions for orientation tuning, measured via circular variance and the modulation ratio, $F_1/F_0$, which is often used to characterize cells as either simple or complex (Mata & Ringach, 2005). Figure 10A shows the distributions for all cortical neurons in the network. The shapes of both distributions, specifically the negatively skewed CV and the bimodal $F_1/F_0$, are consistent with experimental data. Figure 10B shows the distribution of informative neurons for the three
A

Psychometric area under ROC

Psychometric area under ROC

B

Neurometric area under ROC

Correlation coefficient

% Phase coherence

0

0.1

0.2

0.3

0.4

0.5

0.6

0.7

0.8

0.9

1

0

35%

40%

45%

0.2

0.4

0.6

0.8

1

Figure 12. Statistical analysis for establishing the correlation between the fundus image, model prediction, and behavioral data. (A) Scatter plot of the psychometric Az values versus neurometric Az values. There is a significant positive correlation between these two quantities. (B) The absolute correlation coefficient between drusen index and psychometric area under the ROC curve (Az values (white bars) and the absolute correlation between neurometric Az and psychometric Az values (black bars). Asterisk indicates statistically significant difference at \( p < 0.05 \).

drusen cases shown in Figure 7. Note there are no substantial differences between the shape of the distributions (specifically skewness and bimodality) for the full network vs. the informative neurons selected by the decoder for these drusen cases. The decoder thus selects from a broad population of cells in terms of orientation tuning and modulation ratio and is not overly biased toward neurons having specific classical tuning.

We also examined the spatial distribution of neurons selected for this same AMD case, with results shown in Figure 11. We find that neurons that are mapped to the spatial locations of drusen are also selected by the decoder. Thus, though the drive from the retinal input is reduced due to the drusen, these neurons still convey discriminative information due to the recurrent connections and resulting neurodynamics. We interpret this result as a form of “filling-in” that is not just filling in of activity but of activity that is discriminatory and also is purely due to short-range cortical connectivity in the model.

Comparing fundus-derived and neurometric-derived predictions

Statistical analysis was carried out to investigate the relationship between the fundus image, the model prediction, and the behavioral data. We first investigated the predictive value of our model simulations. Figure 12A shows the psychometric Az values (i.e., area under the corresponding ROC curve) versus neurometric Az values, across all the phase coherences. The two are highly correlated (\( p < 0.05 \)), and neurometric Az is an extremely good predictor for psychometric Az (Az\textsubscript{psycho} = 1.006 \times Az\textsubscript{neuro} − 0.0647). We then compared the predictive value of our model to more conventional predictive measures based on direct analysis of the fundus image. To characterize the fundus images, we defined the drusen index (DI) as the fraction of drusen-free area on the fundus:

\[
DI = \frac{\int_{\Omega} \rho^{AMD}(\vec{y}) d\vec{y}}{\int_{\Omega} d\vec{y}},
\]

where \( \Omega \) defines the region within the red square on the retinal mask, while \( \rho^{AMD}(\vec{y}) \) is the binary retinal mask, \( \vec{y} \in \Omega \). We calculated the correlation between this drusen index and psychometric performance, as well as the correlation between neurometric performance and psychometric performance. This was done for each phase coherence. Figure 12B plots the absolute value of the correlation coefficients. The neurometric performance derived from the computational modeling approach correlates better, on average, with the psychometric performance. The largest and most significant difference is for 35% coherence, which in fact is the closest coherence to psychophysical threshold for most subjects in our study.

Discussion

We used a large-scale model of V1 to map the retinal impairment, measured through fundus imaging, onto
cortical activity. The sparse decoder subsequently maps the cortical activity into simulated behavior of the model for a 2-AFC task. The combination of the V1 model and the decoder provides a computational framework to examine the cortical and perceptual consequences of vision loss resulting from AMD.

Our results indicate that though the psychophysics shows the largest differences between control subjects and AMD patients for high SNR stimuli (e.g., high coherences in our paradigm), the greatest predictive power in terms of the correlation of psychometric and neurometric performance is for stimuli at intermediate SNRs—i.e., stimuli near psychometric threshold (e.g., 75% correct). We also observe, through analysis of the spatial distribution of the neurons selected by the decoder, a form of “informational filling-in,” namely, that some neurons that receive minimal input drive nevertheless have significant discriminatory information in their neurodynamics. This type of filling-in is different than “perceptual filling-in,” which often has been described within the context of the retinal scotomas (Zur & Ullman, 2003) and in which the neurons representing features that are absent (e.g., oriented lines/edges) are modulated by contextual inputs. Nonetheless, it is every bit as significant since perception includes not just constructing the scene but also analyzing the scene, whereas discrimination falls into the latter.

We used a decoding framework that imposes a sparsity constraint on the neural representation. This of course is not the only way one could decode the activity. However, our motivation for using this approach is that it is relatively simple, being a linear decoder, and employs a sparsity constraint that is in line with how the early visual system is hypothesized to encode the stimulus. Though our decoding model is meant only to be a method for analyzing the neurodynamics, it is worthwhile to point out that the decoder in effect operates like a simple “grandmother cell” (Gross, 2002). Though the optimization for learning the weights of the decoder is not biologically based, the functional properties of the decoding, including the sigmoidal response function imposed by the logistic mapping, are consistent with single neurons in the ventral stream being selective for complex objects (Quiroga et al., 2005).

It is important to reiterate several assumptions we made for our analysis. First, we treat all drusen as scotoma and model them using binary masks, which is a first-order approximation for the retinal impairment. Second, we only consider dry AMD cases, since wet AMD often involves complicated pathologies such as neovascularization, which causes much more damage to central visual function. Third, we assume that the patients use their fovea as their preferred retinal locus; moreover, our model of V1 only covers 2° × 2° of visual field.

Our cortical model examines the direct link between drusen areas and perceptual performance, assuming that the “scotoma” is the only cause of vision loss. However, the degradation of perceptual vision loss in AMD patients can be caused by multiple factors and not only limited to the drusen. Additional factors/causes of perceptual vision loss could be attributed to retinal pigment epithelium pathology, generalized photoreceptor dysfunction, etc.

Filling-in naturally occurs in the model via the short-range connections, neurodynamics, and receptive field scatter. This is seen in the simulated firing rate patterns in the magnocellular cortex. Of course, other mechanisms for filling-in are likely involved including long-range intracortical connections and extrastriate feedback, as well as the effect of reorganization of receptive fields for ganglion, LGN, and cortical cells.

The computational framework combining a realistic model of V1 and a sparse linear decoder, nevertheless, gives us a quantitative framework to predict the cortical and perceptual consequences of retinal impairment. Since the simulations are personalized to each patient, via their fundus image, the framework potentially provides a quantitative assessment for relating clinical findings in retinal imaging to perceptual function.

Acknowledgments

This work was supported by NGA NURI Grant HM1582-07-1-2002, NIH Grant R01EY015520, and the New York Community Trust. We thank Marios Philiastides for assistance in the statistical analysis and Mihai Busuioc for assistance with the psychophysics experiments.

Commercial relationships: none.
Corresponding author: Paul Sajda.
Email: psajda@columbia.edu.
Address: Department of Biomedical Engineering, Columbia University, 351 Engineering Terrace, MC8904, New York, NY 10027, USA.

References


