Simulation of visual perception and learning with a retinal prosthesis

James R. Golden[1], Cordelia Erickson-Davis[2,6], Nicolas P. Cottaris[3], Nikhil Parthasarathy[1,7], Fred Rieke[4], David H. Brainard[3], Brian A. Wandell[5], E.J. Chichilnisky[1]

[1]Neurosurgery, Ophthalmology & Hansen Experimental Physics Laboratory, Stanford University, Stanford, CA 94305 USA
[2]School of Medicine, Stanford University, Stanford, CA 94305 USA
[3]Psychology, University of Pennsylvania, Philadelphia, PA 19104 USA
[4]Physiology & Biophysics, University of Washington, Seattle, WA 98105 USA
[5]Center for Image Systems Engineering, Stanford University, Stanford, CA 94305 USA
[6]Department of Anthropology, Stanford University, Stanford, CA 94305 USA
[7]Institute for Computational & Mathematical Engineering, Stanford University, Stanford, CA 94305 USA

The nature of artificial vision with a retinal prosthesis, and the degree to which the brain can adapt to the unnatural input from such a device, are poorly understood. Therefore, the development of current and future devices may be aided by theory and simulations that help to infer and understand what patients see. A novel computational framework was developed to predict visual perception and the effect of learning with a subretinal prosthesis. The framework is based on the idea that the central visual system efficiently reconstructs the incident image from the retinal output. To implement this idea, a simulation of the normal responses of the major retinal ganglion cell types was used to deduce the optimal linear reconstruction of the visual stimulus from retinal activity. The result was then used to make inferences about visual experience with simulated retinal activation by a subretinal prosthesis. The inferred visual perception obtained with prosthesis activation was substantially degraded compared to the inferred perception obtained with normal retinal responses, as expected given the limited resolution and lack of cell type specificity of the prosthesis. Consistent with the importance of cell type specificity, reconstruction using only ON cells, and not OFF cells, was substantially more accurate. Finally, when reconstruction was re-optimized for electrical stimulation, simulating learning by the patient, the accuracy of inferred perception with prosthesis stimulation was closer to that of natural vision. The reconstruction approach provides a framework for interpreting patient data in clinical trials, and may be useful for improving prosthesis design.

Correspondence: E.J. Chichilnisky, ej@stanford.edu. Funding: Simons Foundation Collaboration on the Global Brain (EJC,BW,DB,FR), Stanford Neurosciences Institute (EJC,SM), DARPA Contract FA8650-16-1-7657 (EJC). Acknowledgements: We thank Vincent Bismuth and Daniel Palanker for valuable discussions and comments on the manuscript, Nishal Shah for technical assistance and useful input, and Georges Goetz for helpful discussions.

Introduction

Retinal prostheses can restore visual perception to blind patients through electrical stimulation of surviving neurons in the retina, based on images captured by a camera (Humayun et al. 2012; Zrenner et al. 2011). The perceptual experience of patients with existing devices, however, is far from that of a person with normal vision. Several efforts are underway to build next generation devices (Mathieson et al. 2012; Lorach et al. 2015a), and clinical experience with other types of neural implants suggests that learning by the patient may help to overcome some of the technical limitations of current devices (Hallberg and Ringdahl 2004; Shannon 2012; Eisen 2003; Waltzman et al. 1993). Nonetheless, the degree to which technical improvements and learning can improve patient outcomes remains poorly understood, and current understanding of the experience of artificial vision itself is limited. Thus, it would be valuable to have a framework to predict the visual experience that retinal prostheses will provide to patients, as well as how that visual experience may improve with learning. Such a framework could help researchers and clinicians to interpret what patients see, and could guide future technical development based on understanding the factors that limit performance.

The goal of this work is to provide such a framework. The challenge is to develop a simulation of prosthesis function that goes beyond empirical observations and features of the device to produce quantitative predictions of visual experience, grounded in our understanding of the human visual system. To approach this problem, a key conceptual step is proposed: the assumption that perceptual experience is an attempt by the visual system to reconstruct the incident visual image from the retinal signal. With this simple but powerful assumption, a simulation of visual perception and learning with a prosthesis can be accomplished as follows.

First, the optimal linear reconstruction of the image from retinal activity, deduced from simulations of retinal responses to natural scenes, is computed to produce a prediction of perception for any given evoked pattern of retinal activity (Nirenberg and Pandarinath 2012). Linear reconstruction can be understood as placing a particular spatio-temporal filter in the reconstructed image for each spike produced by a given retinal neuron, and summing the contributions of all neurons (Warland et al. 1997, Rieke 1997). A unique, optimized reconstruction filter is used for each neuron, reflecting the information that neuron conveys about the image in normal vision. This approach, coupled with a simulation of how the device activates retinal neurons, produces a specific prediction of the perceived image with a retinal prosthesis. Second, the potential impact of a patient learning on artificial vision is then inferred by recomputing the filters for each neuron on the basis of prosthesis stimulation, rather than on the basis of normal visual responses. Thus, the framework permits the evaluation of prosthetic vision compared to normal vision, as well as the effect of learning.

Here we develop this approach using simulations of healthy retinal function (ISETBio; Brainard et al. 2015; Jiang et al. 2017) and simulations of retinal activity evoked by a photovoltaic subretinal prosthesis (Mathieson et al. 2012; Wang et al. 2012; Lorach et al. 2015a) on the basis of data on isolated retina (Goetz et al. 2015; Mandel et al. 2013; Boinagrov et al. 2014; Lorach et al. 2015b). The approach yields simple quantitative predictions of two aspects of prosthetic vision. First, the predicted image obtained with a retinal prosthesis is degraded compared to the image obtained in normal vision over the same field of view. This finding is consistent with the limited resolution and cell type specificity of prostheses. Second, re-optimizing filters for prosthesis stimulation substantially improves the reconstructed image. This indicates that the potential for learning by patients to improve prosthesis function is high. We comment on potential improvements in the approach, on the use of this framework for design of future prosthesis technology, and on the potential for image processing to optimize the function of existing devices.

Methods

Overview

A computational framework was developed to infer a visual percept based on RGC activity, both for healthy light activation, and for electrical activation using a subretinal prosthesis. The ISETBio (Image Systems Engineering Toolbox – Biology) software was used to model transformation of a visual stimulus into RGC spikes in the healthy retina (Fig. 1). Additions to software were used to simulate electrical activation of bipolar cells via a retinal prosthesis, in turn producing RGC spikes. A linear transformation of simulated RGC responses was then used to infer the perceived image, over samples of a large image data set. The linear transformation was selected to optimize reconstruction for either the healthy or prosthesis condition.

Healthy Retina Simulation

The ISETBio simulation begins with a representation of image radiance. This is transformed into retinal irradiance via the simulated cornea, lens, and inert pigments. Irradiance is then transformed into a spatial array of photon absorptions and photocurrent for each cone in the photoreceptor mosaic. Cone photocurrents are then linearly transformed into bipolar responses, with appropriate cone type connectivity. Finally, bipolar responses are transformed into RGC responses via a generalized linear model that produces simulated spikes. This package simulates many aspects of healthy retinal responses to arbitrary visual stimuli, although it does not yet

explicitly include the role of horizontal and amacrine cells or diverse nonlinear aspects of retinal processing. Below, the full spatio-temporal model is described. Later, the more limited use of the model in simulating static image presentations, RGC responses, and reconstruction is presented.

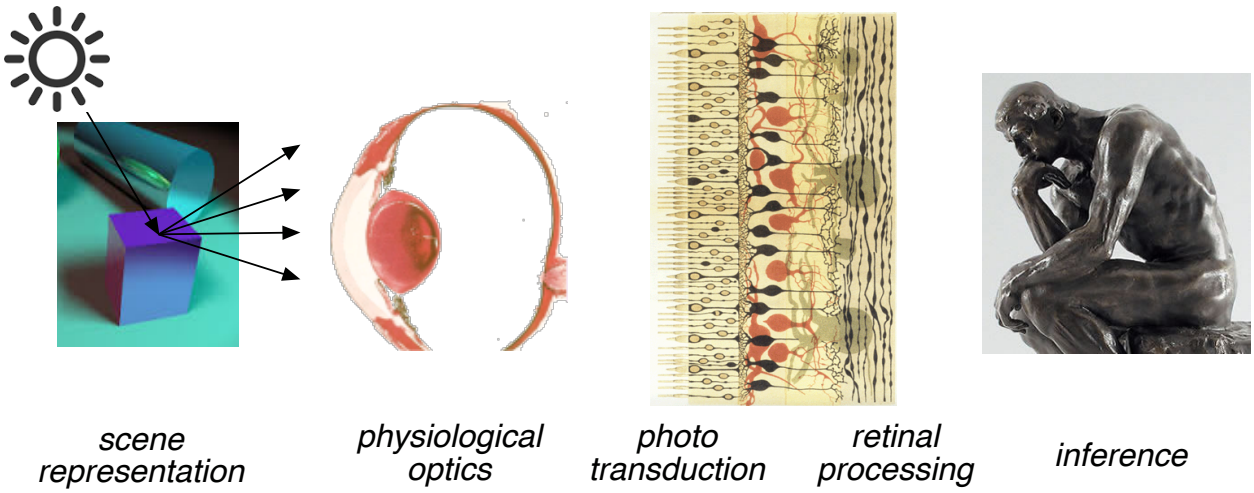


Figure 1. The ISETBio computational pipeline. Stimuli represented by spectral radiance over space were transformed to retinal irradiance, cone photopigment isomerizations, photocurrent, bipolar responses, and finally RGC spikes.

The stimulus image was modeled as an array of MxN pixels on a RGB computer display, and summarized by an MxNx3 matrix. The spectral power density of the display primaries was used to calculate a larger array (MxNxW) representing the display radiance, where each value of W indexes the radiance at a particular wavelength. The full spectral image was then passed through a model of the human optics based on mean wavefront aberrations measured on-axis (near fovea) as well as lens and macular pigment densities. The retinal spectral irradiances were used to compute the Poisson distributed photopigment isomerization rates for the L, M and S cones in the simulated cone mosaic. Cone isomerizations were computed using photopigment parameters for human foveal vision, which reproduce the CIE cone fundamentals (CIE 2006). A linear temporal filter was applied to the cone isomerizations to generate simulated cone photocurrents over time (Angueyra and Rieke 2013).

Four types of bipolar cells were modeled, ON diffuse, OFF diffuse, ON midget, and OFF midget. Each type pooled inputs linearly over a local collection of cones. The bipolar temporal response was assumed to be equal to that of the RGC type to which it provided input (see below). RGC temporal responses were obtained from a standard data set (Pillow et al. 2008). Thus, the bipolar temporal response was obtained by convolving the cone response in time with a filter obtained

by deconvolving the RGC temporal impulse response by the cone temporal impulse response, for the appropriate RGC type. The connectivity of the three cones types to the four types of bipolars simulated was modeled as follows (Dacey 2000): the first three types of bipolar cells received input from only L and M cones, the fourth received weighted input from S cones (25%) in addition to input from L and M cones (Field et al. 2010).

The RGC response was simulated with a generalized linear model based on bipolar cell input (Pillow et al. 2008). The four bipolar types above provided sole input to ON parasol, OFF parasol, ON midget and OFF midget RGCs, respectively. The input to the RGC was computed as a local sum of bipolar cell responses, chosen so that the stimulus-referred spatial properties of the RGC fitted a published data set (Pillow et al. 2008). Because the bipolar temporal response was already chosen to equal that of the RGC, the transformation from bipolar to RGC did not involve further temporal filtering. The input to the RGC from bipolars was then followed by exponentiation and Poisson spiking with a post-spike filter, obtained from fits to experimental data (Pillow et al. 2008).

The simulated patch of retina was a square region 1.7° (512 µm) on a side, centered at 1.8° eccentricity. The simulated patch had a uniform density of cones, bipolars, and RGCs appropriate for 1.8° eccentricity, and did not account for the changing density of these cells over the field of view. The cone mosaic over the region was a 256x256 square lattice with a pitch of 2 µm. The mosaics of each bipolar cell type had the same dimensions as the cone lattice. Each bipolar cell integrated over cone inputs using a difference of Gaussians spatial weighting function with center peak amplitude of 1, surround peak amplitude of 0.01, center SD of 4 µm, 4 µm, 2 µm and 2 µm for the four types of bipolar cells respectively, and surround SD equal to 9 times the center SD (Dacey 2000). The RGCs in the simulation consisted of approximately hexagonal lattices of 28x32, 31x35, 55x63, and 61x70 cells for the four RGC types, respectively, yielding a total of 9,716 RGCs. The four RGC types received inputs from nearby bipolars, weighted according to a difference of Gaussians function with center peak amplitude of 1, surround peak amplitude of 0.375, 0.375, 0.5 and 0.5 for the four RGC types, center SD of 10 µm, 8 µm, 4 µm, and 3.5 µm, and surround SD equal to 1.15 times the center SD. These parameters were selected to approximately match experimental data (Pillow et al. 2008; Croner and Kaplan 1995). Poisson noise in photoisomerization and Poisson spiking by RGCs were the only sources of noise in the simulation.

Subretinal Prosthesis Simulation

To model the encoding of the visual signal by a subretinal prosthesis, a simulation of the electrical stimulation of bipolar cells was introduced into the ISETBio pipeline. The simulation

was modeled after a subretinal prosthesis, in which a visual stimulus captured by a camera mounted on goggles is transformed into pulses of infrared light that irradiate a grid of photovoltaic pixels implanted in the sub-retinal space near the bipolar cells (Mathieson et al. 2012). In the simulation, an 8x8 rectangular array of photodiodes with 70 µm pitch covered the field of view. The voltage on each photodiode at each time step was assumed to be proportional to the irradiance of the visual stimulus integrated over the photodiode area during that time step. The electrodes produced responses in nearby bipolar cells in proportion to this voltage. The spatial profile of bipolar activation was the profile of the grid of electrode center points, each scaled according to stimulation voltage, convolved with a Gaussian function with SD=35 µm. Each bipolar cell was activated in proportion to the value of this activation profile at its location. The constant of proportionality was set to approximately mimic the distribution of RGC responses that occurs with natural scene stimulation (data not shown). With this constant, the stimulus contrast required for reliable detection of a full-field step of light (see below) was 3.1%. The temporal response of the bipolar was given by the activation value convolved with the bipolar temporal impulse response. This implementation did not capture nonlinear effects such as saturation of the photovoltaic pixels at high input irradiances (Mathieson et al. 2012).

Static Simulation

Although the calculations above yield simulations of RGC activity over time, most of the analysis was performed for static images and static RGC responses, for simplicity. In both the healthy retina and prosthesis simulation, images were presented for 0.4 s, and the responses of RGCs were accumulated over the time window 0.320-0.328 s after the stimulus presentation. This produced a vector of spike counts across the RGC population, in response to each static image.

Optimal Linear Reconstruction

Linear reconstruction was used to infer the image perceived by a patient based on RGC activation. To perform linear reconstruction, a spatial filter was first identified for each RGC. Each spike emitted by that RGC during the response caused a copy of the spatial filter to be placed in the reconstructed image. The final reconstructed image was the sum of the filters contributed by all spikes from all RGCs. The spatial filters for the collection of RGCs were selected so that, when applied to RGC responses in either healthy or prosthesis simulation, produced reconstructed images with least squared deviation of pixel intensity from the original stimulus (see below). The identification of optimal filters was performed on 80% of the images in the data set. The performance of reconstruction was then assessed on the remaining 20% of

images. In total, 288,000 image patches were used, each 100x100 pixels, sampled from the Imagenet database (Deng et al. 2009).

Reconstruction filters were calculated using linear regression (Warland and Meister 1997). Let s denote the stimulus image as a row vector, let r denote a row vector indicating the number of spikes from each RGC with an offset term of 1 as its first entry, and let W denote the reconstruction matrix, in which each row contains the reconstruction filter for a particular RGC. The reconstruction s_{rec} of the stimulus s is then given by:

$$s_{rec} = r\,W$$

To obtain a matrix W that reconstructs as accurately as possible for a collection of images, let S be a matrix containing an image in each row, let R be the matrix containing the corresponding responses from all cells in each row, and let S_{est} be a matrix containing the reconstructed images in each row. The objective is to minimize the squared difference between reconstructed and original images,

$$\| S_{rec} - S \| = \| R\,W - S \|$$

Where $\|\bullet\|$ represents the sum of squared entries. This is a standard regression problem with the solution (Menke 2012):

$$W = (R^T R)^{-1} R^T S$$

The performance of reconstruction using the matrix W was then assessed on a set of held out stimuli and responses. As expected, reconstruction performance was limited by the amount of data used to identify the entries of W. To obtain robust estimates of W, regularization was performed by limiting the rank of W (Chambers 1977). Specifically, if R is given by its singular value decomposition,

$$R = U\,D\,V^T$$

Then, from the expression for W above, we have:

$$W = V\,D^{-1}\,U^T\,S$$

Thus, an approximation to W with rank K is given by:

$$W_K = V D^{-1}_K U^T S$$

Here D_K denotes zeroing all but the first K diagonal entries of D (singular values). For the healthy retina reconstruction, retaining the top 50% of singular values (K=4904) produced optimal performance. For the prosthesis reconstruction, retaining the top 5% of singular values (K=491) produced optimal performance. These values were used in all analyses above.

Linear Classification for Comparison to Clinical Tests

In order to simulate measurements of perceptual thresholds, such as those performed on patients with previous retinal implants (Stingl et al. 2013), a linear classification approach was used. Two sets of stimuli were processed by the prosthesis simulation, producing reconstructed images, and a linear classifier was trained to discriminate pairs of reconstructed images, one from each set. Classification performance was then assessed on held out data as a function of a particular stimulus variable, e.g. contrast or spatial frequency. Discrimination thresholds were defined as the value at which the classifier achieved 82% accuracy.

Results

To quantify normal visual system function, as a benchmark for artificial vision with a prosthesis, a simulation of the normal (healthy) light activation of retinal ganglion cells (RGCs) by visual images was produced. This simulation was performed using the publicly-available ISETBio software (see Methods). Briefly, each RGC was modeled as a linear-nonlinear cascade, with center-surround receptive field organization (Kuffler 1953). Simple cascade models have been used to capture the responses of populations of RGCs in previous studies (Chichilnisky 2001, Chichilnisky and Kalmar 2002, Field and Chichilnisky 2007, Pillow et al. 2008). The simulation was built on modeled inputs from photoreceptors transmitted through bipolar cells in the retinal network. The simulation was focused on the four numerically dominant RGC types in macaque and human retinas: ON parasol, OFF parasol, ON midget, and OFF midget (Dacey 1999). Each of the four cell types forms a mosaic uniformly covering the visual scene (Fig. 2, top). The spatial, temporal, nonlinearity, and noise properties of each RGC type (Fig. 2) were taken from experimental findings and incorporated into the software. Presentation of a static natural visual image (Fig. 3, left) to the simulated retina causes a spatial pattern of activation in the four RGC types (Fig. 3, second and third columns). Although the simulation produced RGC spikes over time, for simplicity most of what follows focuses on flashed static images reconstructed from RGC spike counts (see Methods).

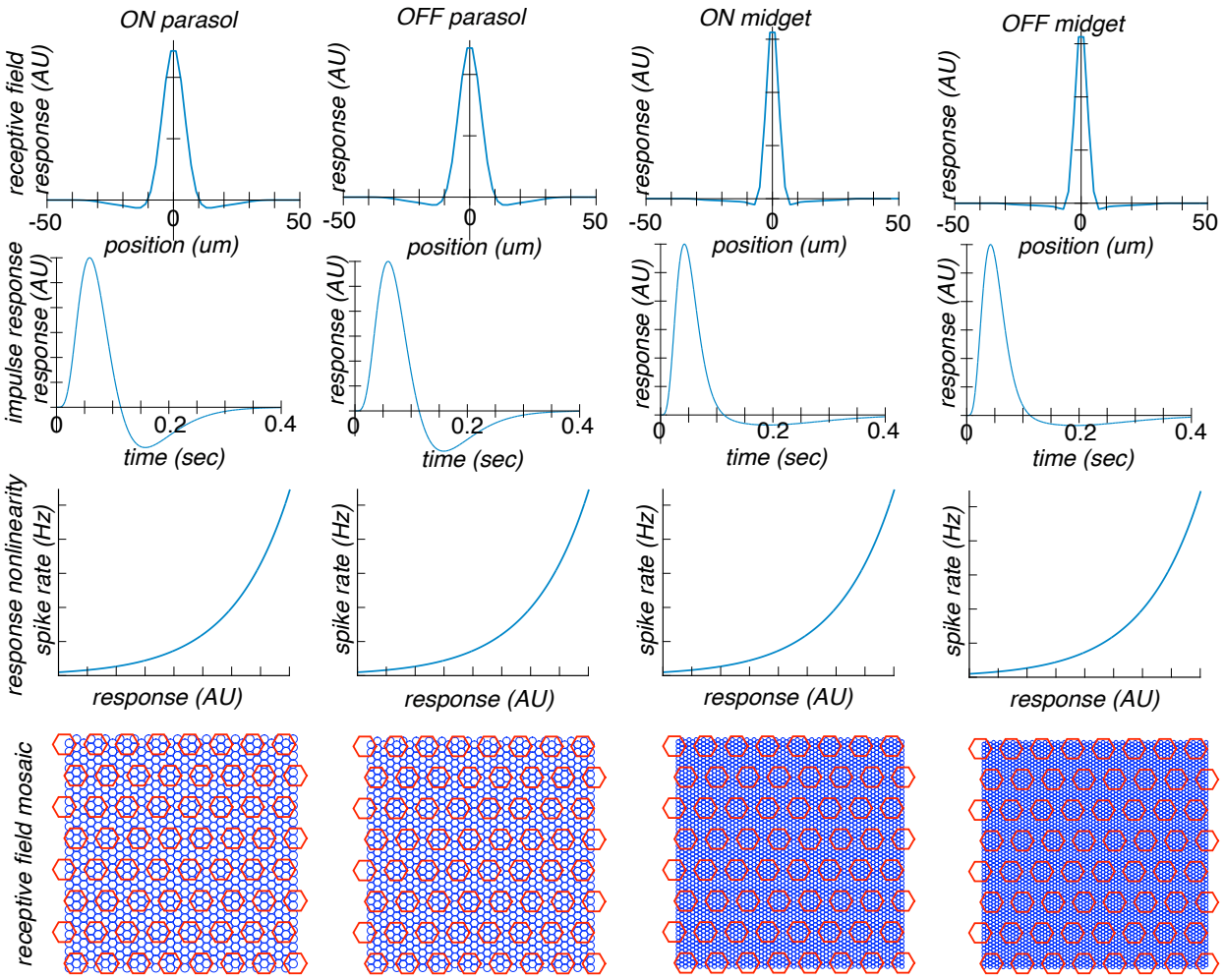


Figure 2. Simulation of RGC responses. Each RGC is modeled as a linear-nonlinear cascade (see Chichilnisky 2001, Chichilnisky and Kalmar 2002), with an intermediate representation of photoreceptor and bipolar cells (see Methods). Top row: cross-sectional spatial profiles of the spatial receptive fields of each RGC type, with midget cell RF extents smaller than parasol, and OFF slightly smaller than ON. Second row: the temporal impulse response of each RGC type; a typical biphasic response, with different details in midget and parasol RGCs. Third row: the response nonlinearity of each RGC type is an accelerating half-wave rectifying function. Bottom row: the collection of RFs of each RGC type (blue circles) uniformly tile the region of retina, with different densities, and the implant electrodes (red circles) form a much coarser sampling of space.

To simulate the retinal signal that would be produced by a prosthesis, a model was developed of the activation of the retinal network by the electrodes of a wireless subretinal photovoltaic prosthesis (Loudin et al. 2007). The prosthesis technology has been tested extensively in isolated retina as well as *in vivo* animal studies (Goetz et al. 2015; Mandel et al. 2013; Boinagrov et al. 2014; Lorach et al. 2015b). In the simulation, the presentation of a static visual image to the

prosthesis produced spatial patterns of current from the electrodes. This current then caused depolarization of the various bipolar cell types that provide input to the ON and OFF parasol and midget RGCs, resulting in spiking activity in the corresponding RGC populations (Fig. 4, second and third columns). The pattern of activation induced by the prosthesis was quite different from the activation of RGCs in the healthy retina simulation (Fig. 3, second and third columns). A salient difference is that the normal opposite activation of ON and OFF populations by light and dark portions of the visual scene was not reproduced, because light-evoked current in the prosthesis activates ON and OFF bipolar cells indiscriminately and with the same polarity. This lack of cell type specificity is a known limitation of existing prosthesis technology (Goetz et al. 2015; Goetz and Palanker 2016; Jepson et al. 2013, 2014).

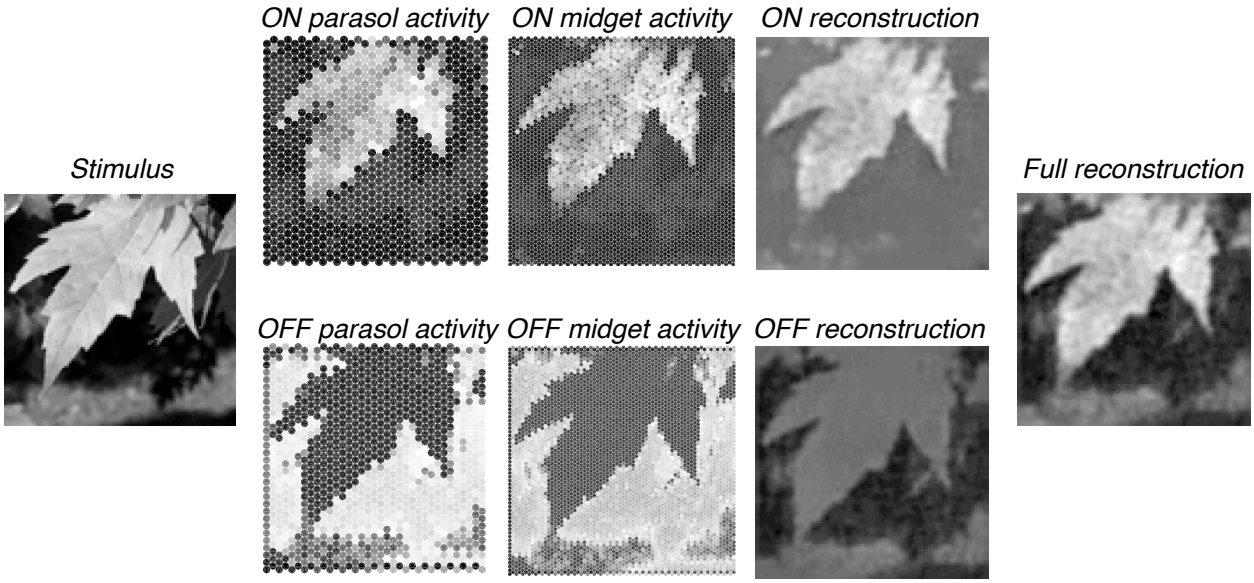


Figure 3. Simulation of visual stimulus, response, and linear reconstruction in healthy retina. Left: a sample visual stimulus. Second and third columns: simulated responses of the four major RGC types over space to the flashed stimulus, with light colors indicating higher activity. Fourth column: linear reconstruction of the visual input by ON cells and OFF cells, using the responses shown in the second and third columns, and reconstruction filters trained on a large collection of responses to natural images. Right: linear reconstruction of the original stimulus from all four RGC types combined.

To assess how the different patterns of RGC activity with normal and prosthesis stimulation may influence visual perception, a simulation of the perceived image based on activation of RGCs was produced. The key assumption of the simulation is that the visual system reconstructs the visual image from the normal pattern of retinal activity in a way that is optimal for healthy vision, producing a reconstructed image that resembles the real visual scene as closely as possible. Specifically, in the simulation, patterns of RGC activation are linearly transformed into

simulated "perceived" images that deviate as little as possible from the original images in terms of mean squared error, with the mean taken across an ensemble of natural images (Warland 1997, Rieke 1997, Nirenberg and Pandarinath 2012). These key technical assumptions (linear reconstruction, squared error, natural images) are considered further in Discussion. Based on these assumptions, image reconstruction is a regression problem, in which an optimal linear transformation is calculated that transforms RGC activation to the incident image, over 230,400 natural images (see Methods), and the transformation is then evaluated on 57,600 held-out images. Conceptually, this linear transform amounts to assigning a fixed spatial filter to each RGC (Fig. 5, left), such that each spike from the RGC contributes that filter to the reconstructed image, and contributions from all RGCs are summed. The filter for each cell can be understood as an interpretation of the visual information conveyed by a spike in that cell.

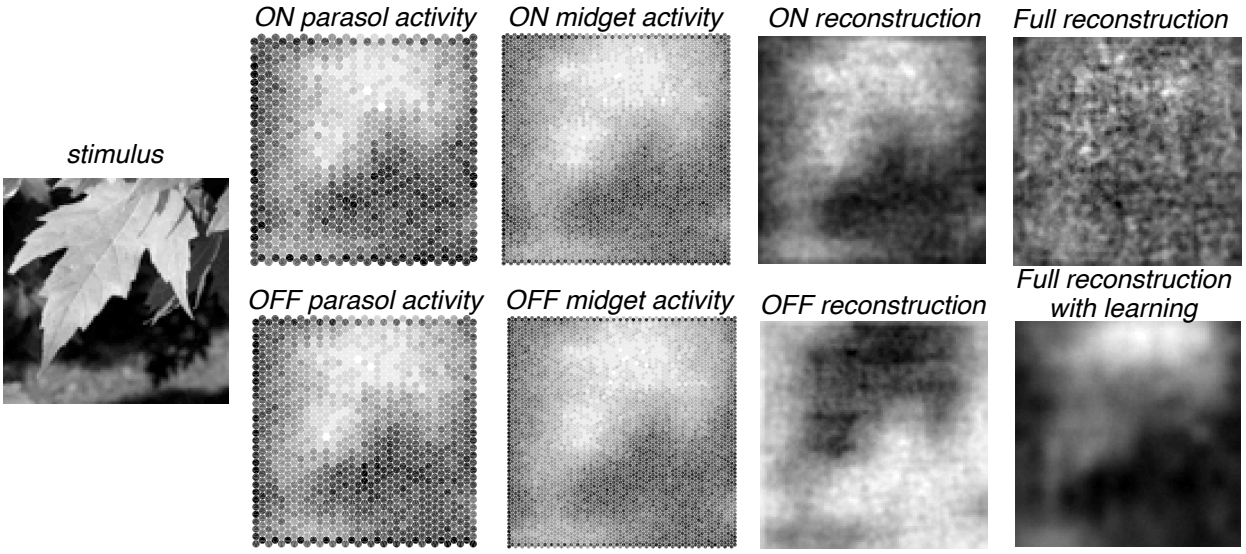


Figure 4. Simulation of stimulus, response, and linear reconstruction with retinal prosthesis. Left: visual stimulus. Second and third columns: simulated responses of four major RGC types to stimulation with the prosthesis. Note the similar patterns of activation of ON and OFF cells by the prosthesis. Fourth column: linear reconstruction of the stimulus, using the reconstruction filters obtained from the healthy retina simulation, separately using ON and OFF cells. Note the reverse polarity of the reconstruction for OFF cells. Right, top: reconstruction with all four RGC types combined normally, causing cancellation of ON and OFF signals. Right, bottom: reconstruction with all four RGC types combined, using reconstruction filters obtained with prosthesis stimulation, mimicking optimal learning by the patient.

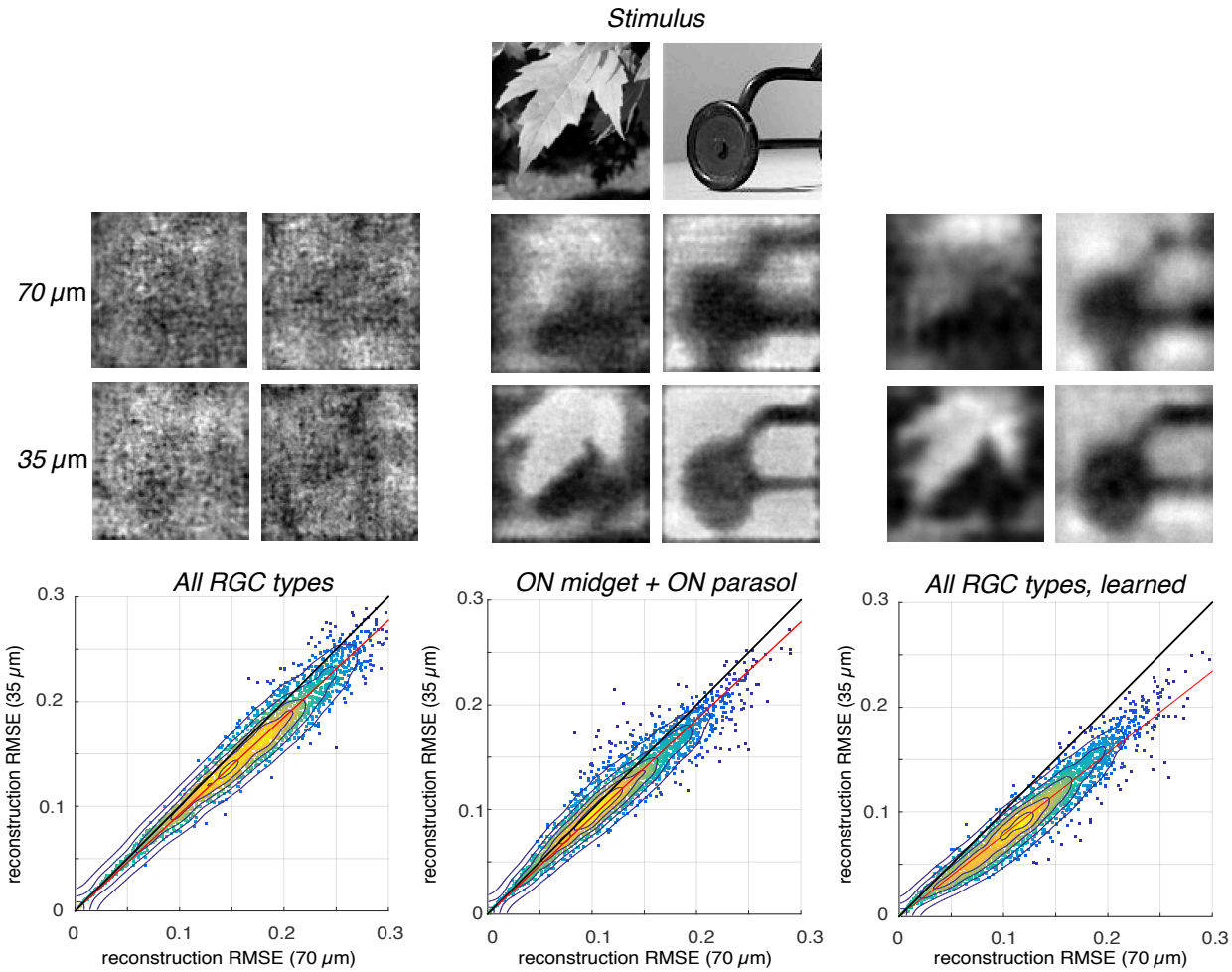
In the case of natural activation of RGCs, linear reconstruction produces an image that resembles the original image (Fig. 3, right), with a resolution that is limited by the density of RGCs. Because the reconstruction is linear, the contributions of the different RGC types can be

examined separately. For example, ON midget and ON parasol cells provide similarly detailed contributions the reconstruction as OFF midget and OFF parasol cells (Fig. 3, fourth column).

In the case of prosthetic activation of RGCs, the naive reconstruction procedure that was optimized for normal vision produces a degraded image, in which few features of the original image are easily recognized (Fig. 4, top right). This degradation is not surprising given that the ON and OFF bipolar cells, and correspondingly ON and OFF RGCs, are activated similarly by the prosthesis model (Fig. 4, second and third columns), while spikes from ON and OFF RGCs are treated oppositely in normal reconstruction (Fig. 6, left) – thus, their contributions would often be expected to cancel. This interpretation was confirmed by examination of the ON and OFF reconstructions separately: the two groups of cells each produced more accurate reconstructions (ignoring the sign of the reconstructed contrast) individually (Fig. 4, fourth `column) than when combined. Averaged over test images, the root mean square (RMS) error for reconstruction with all cells was 0.09 in the normal retina simulation, and 0.17 for the prosthesis simulation. As a comparison, the RMS error obtained with temporally shuffled spikes bearing no relationship to the visual stimulus was 0.24. Thus, a computational framework that assumes ON and OFF RGC signals are equally stimulated and used to reconstruct the visual scene, in a way that is optimal for normal vision, predicts that the perceived visual image with a prosthesis is highly degraded. Under these conditions, factors that might otherwise be expected to improve prosthesis function, namely the density of electrodes and magnitude of current spread, had little impact. Specifically, reducing electrode spacing and current spread by a factor of two produced a decrease in RMS error of only 7% (Fig. 5, left).

The naive prediction of severely degraded artificial vision above appears inconsistent with patient reports, which indicate that some degree of spatial vision is in fact possible with a prosthesis (see Goetz and Palanker 2016). One possibility, suggested by the dominance of "bright" (rather than "dark") phosphenes reported by patients (Stingl et al. 2013), as well as by models showing asymmetries in electrical activation of ON vs. OFF cells (Werginz et al. 2015; Cottaris and Elfar 2005), is that the contribution of ON cells to artificial vision is stronger than that of OFF cells, reducing the cancellation of visual signals. To simulate this hypothesis in an extreme form, reconstruction using only ON midget and parasol cells was examined (Fig. 4, top row, fourth column) with no contribution from OFF midget and parasol cells. In this simulation, the average RMS reconstruction error across images was 0.14, a value higher than in the healthy condition, but lower than the error obtained with prosthesis stimulation and reconstruction with all four RGC types. In this simulation, the additional improvement in performance with higher-resolution stimulation was still 7% (Fig. 5, middle). In sum, the artificial vision provided by

retinal prostheses in the clinic could benefit substantially from a dominance of specific cell types



in artificial vision.

Figure 5. Reconstruction performance with denser electrodes and lower current spread. Lower panels shows the root mean square error (RMSE) for reconstruction of 57,600 tested images, using the original electrode spacing and current spread (70 µm, 35 µm) vs. values obtained with both reduced two-fold (35 µm, 17.5 µm). Left: reconstruction error using all four RGC types. Red line shows linear regression to the data; regression slope 0.93. Middle: reconstruction error using only ON cells; regression slope 0.93. Right: reconstruction error using all four RGC types, with learning; regression slope 0.78. Upper panels show example reconstructed images from each set of parameters.

Could artificial vision improve if the brain learns to interpret the unnatural patterns of activation of RGCs that occur with a prosthesis? To answer this question, optimal reconstruction filters were recalculated using RGC activity elicited by the prosthesis, instead of natural RGC activity. This simulation of learning mimics a central visual system that learns optimally. As expected, the reconstruction filters associated with RGCs were very different in the learning simulation (Fig.

6), because the RGC activity used for learning was very different from the healthy RGC response. Notably, reconstruction filters exhibited asymmetric shapes that varied between cells, and OFF cells often exhibited filters with ON polarity, consistent with the polarity of prosthesis stimulation. With learned filters, reconstruction of images using all RGC types was substantially improved (Fig. 4, bottom right). The RMS error of reconstruction across test images was 0.12, closer to the value achieved in natural vision. In this condition, improvements in electrode resolution produced a greater improvement than in the previous conditions: a two-fold reduction in electrode spacing and current spread reduced RMS error by 22% (Fig. 5, right). Thus, the framework predicts that the potential for patients to learn the unnatural neural code and improve prosthesis function is significant, and if such learning does occur, it enhances the beneficial effects of technical improvements in the device.

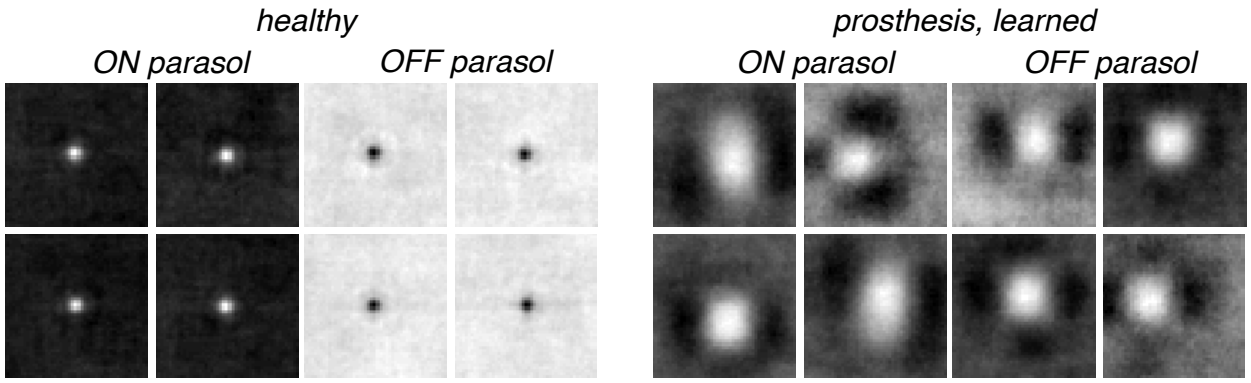


Figure 6. Reconstruction filters in healthy and prosthesis simulations. Left: sample reconstruction filters are shown for ON and OFF parasol cells (four each) obtained from the healthy retina simulation. The filters roughly resemble the spatial receptive fields of the cells (not shown), and their center polarity corresponds to the polarity of light response. Right: sample spatial reconstruction filters for the same cells obtained from the prosthesis simulation. In this case, reconstructions filters are asymmetric and variable, and largely reflect the ON polarity of prosthesis stimulation for both cell types. Images are centered on cell location.

A practical application of this framework to human patients is the prediction of visual performance on standardized tasks, including discrimination of Landolt C and grating orientation. To predict performance on this type of task, visual images were presented to the simulation, and reconstruction was performed as above. Then, the reconstructed images were used to classify the stimulus, using a linear classifier trained on reconstructions obtained from many stimulus presentations (see Methods). For example, in the Landolt C task, stimuli containing a gap were discriminated from stimuli not containing a gap. As expected, the predicted discrimination performance progressively increased with stimulus contrast (Fig. 7).

Discrimination thresholds were roughly ten-fold higher in the prosthesis simulation (with learning) than in the healthy retina simulation. Note that the healthy grating orientation discrimination thresholds were unrealistically low compared to normal human performance, reflecting a classifier that has prior "knowledge" of the stimulus pattern and location, as well as the fact that our simulations capture only some of the noise introduced in the retina, and none of the noise introduced downstream of the retina. Thus, a meaningful quantitative prediction of patient performance would require a more realistic simulation of behavioral discrimination, including consideration of the size of the stimulus, the task, the nature of the classifier applied to the simulated output, and assumptions about noise in the retina and the central visual system. However, using this simplified classification approach, the framework can generate predictions of patient performance relative to normal performance in clinical tasks.

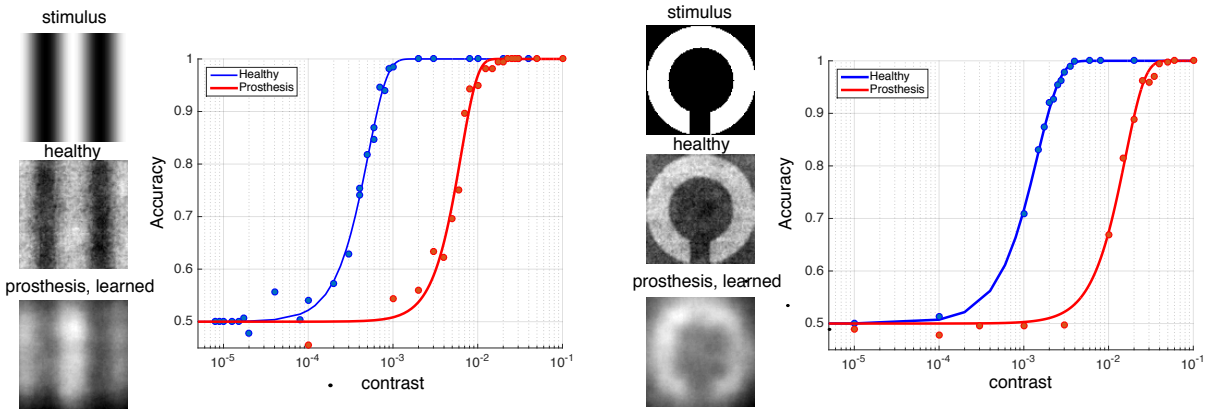


Figure 7. Landolt C and grating orientation discrimination, using healthy and prosthesis simulations. Left: Landolt C stimuli, and similar stimuli lacking a gap, were presented to the simulation, producing reconstructions for healthy retina and prosthesis-stimulated retina with learning. Accuracy of linear classification of stimuli with and without gaps increased with stimulus contrast; classification was nearly an order of magnitude more sensitive with the healthy retina simulation. Right: Grating stimuli, horizontally or vertically oriented, were presented to the simulation and discriminated. A large discrepancy between healthy and prosthesis performance was again observed. Note the unrealistically low normal contrast thresholds, reflecting an highly-trained linear classifier well aligned with the visual stimulus (see Methods).

The above simulations were all performed using only static spatial images and integrated RGC responses, ignoring time. However, the linear reconstruction framework extends naturally to time-varying stimuli and RGC responses, and reconstructed image sequences. In a dynamic reconstruction, a given spike from a given RGC adds a specific spatio-temporal filter to the reconstruction, aligned in time with the spike. These contributions are summed over cells and

over time. Although a quantitative assessment of performance and learning was not performed, dynamic reconstruction produced recognizable estimates of visual perception over time in both the natural and prosthesis simulation (see supplemental movies). The dynamic reconstructions revealed persistent temporal structure in natural scenes which could improve performance – these temporal correlations are not captured by the static simulations above.

Discussion

A simulation framework was developed to explore visual perception and learning with a retinal prosthesis. The framework produced three primary predictions. First, as expected, artificial vision mediated by a prosthesis may be limited by the indiscriminate activation of ON and OFF cells, and simulated dominance of ON cells improves reconstruction. Second, learning by the patient could produce substantial improvements in the perceived image, in part by compensating for the lack of cell type specificity. Finally, learning could enable further benefits from technology enhancement. Below, we discuss the assumptions underlying the approach, its utility for future technology development, and major caveats.

Impact of indiscriminate cellular activation. In healthy vision, roughly 20 different RGC types communicate different visual information to distinct targets in the brain, via their unique temporal patterns of activity (see Gollisch et al. 2010; Field and Chichilnisky 2007). Therefore, indiscriminate activation of different RGC types by a prosthesis produces a highly artificial retinal code, potentially limiting the quality of artificial vision. The most obvious example is simultaneous activation of ON and OFF cells, which may produce conflicting visual signals in the brain. Although this issue is well known, its impact on visual perception has been unclear. The reconstruction approach provides a framework for predicting and interpreting the effect. For instance, a naive simulation in which ON and OFF cells respond equally to prosthesis activation, and their contributions to reconstruction are optimized for natural vision, predicts that artificial vision will be highly degraded. However, previous work suggests that either ON or OFF cells may respond more strongly to electrical stimulation (Werginz et al. 2015; Cottaris and Elfar 2005), and patient reports indicate that "bright" (rather than "dark") phosphenes dominate increasingly over time (Stingl et al. 2013). A simple simulation of ON dominance – a linear reconstruction in which OFF cells are not used – revealed that it could substantially improve artificial vision. In principle, to obtain a quantitative simulation of ON dominance, patient performance in visual tasks (e.g. detection of increments or decrements) could be used to estimate a weighting factor between 0 and 1 applied to the OFF cell contribution. The reconstruction framework may also be useful for assessing the value of developing retinal

prostheses that can selectively stimulate distinct RGC types (Jepson et al. 2013, 2014, Nirenberg and Pandarinath 2012).

Impact of learning. A closely related issue is the impact of learning by the patient, that is, changes in how the brain processes the retinal signal over time in the presence of retinal prosthesis stimulation. Although it is clear that learning occurs and improves artificial vision to some degree (Stingl et al. 2015; da Cruz et al. 2016), its precise impact and potential to improve artificial vision is unclear. The reconstruction approach provides a framework for predicting and interpreting the effects of learning. In a simulation of ideal learning, linear reconstruction optimized for prosthesis stimulation of the retina produced much more accurate estimates of the visual image, compensating to a substantial degree for the indiscriminate activation of ON and OFF cells. In principle, to obtain a more quantitative account of learning, patient performance on visual tasks could be used to estimate a value between 0 and 1 indicating the degree of learning, that is, the relative contribution of the reconstruction filters optimized for prosthesis vs. healthy conditions. This approach could make it possible to objectively assess the improvements in performance over time in each patient. Note that this approach assumes that the learning process involves enhanced perception, rather than behavioral strategies to deal with degraded perception in specific tasks. The choice of measurement of patient performance would be essential for distinguishing these forms of learning.

Assumptions: linearity, error measure, natural images. Three primary assumptions were used to produce the simulation. First, the brain reconstructs the visual image linearly from RGC responses. Second, this linear reconstruction minimizes the squared error between the perceived and true pixel intensities. Third, reconstruction is optimized for natural images, rather than other image ensembles. All of these assumptions are likely to be incorrect in detail and can be improved upon. Accepting the assumption that the brain reconstructs the visual image (see below), linear reconstruction is the simplest to implement, but may not perform as accurately as more complex approaches. Nonlinear reconstruction is a large domain of inquiry, and improvements are already being made using artificial neural networks (Parthasarathy et al. 2017). Similarly, the definition of optimality in terms of squared error between pixel intensities was chosen for technical convenience. One might expect the visual system to instead emphasize accurate reconstruction of certain spatial scales, or specific image features, or even more generally to produce perceptual representations optimized not for physical accuracy but rather to support effective perceptual decisions in the natural environment (Hoffman and Singh, 2012). The image reconstruction assumption, however, is a pragmatic choice that allows us to make progress. In this spirit, a natural extension of the current approach would be to optimize reconstruction using perceptual image metrics, such as SSIM (Wang et al. 2004). Finally, the use of natural images is predicated on the idea that the visual system evolved to encode the structure

in the visual environment. This assumption is important because the reconstruction filters are shaped by the statistics of the visual inputs. However, the statistics of natural images are incompletely understood, and the particular statistical features which are most important to human evolution are even less clear. For example, human faces may be of high importance, but they represent only a tiny subset of the collection of images used. This problem could potentially be explored by using more specific image databases (Parthasarathy et al. 2017). In general, it is worth noting that the ability to predict patient experience is based on strong linking assumptions that may not hold. Experimental tests of these assumptions represent an interesting future direction. Even if ultimately the assumptions turns out to be too strong, the reconstruction analysis presented here can be modified (e.g. by allowing for variable weighting of ON and OFF inputs) to provide insight about the information carried through a prosthesis to the brain.

Related approaches. Most previous attempts to simulate the perceptual experience of patients with retinal prostheses have used a simple "scoreboard model" of phosphenes, in which pixels of an image are translated into perceived dots of light according to the layout of electrodes (Chen et al. 2009). A notable exception is a recent simulation framework that takes into account different electrical sensitivities of bipolar cells, RGCs somas, and RGC axons (Beyeler et al. 2017). This work reproduces clinical data and accounts for the perceptual effects of axon activation, temporal desensitization, and the distinct effects of pulse frequency and amplitude on phosphene attributes. However, the emphasis is quite different from the current work. Specifically, it is based on an empirical framework rather than a theoretical foundation such as stimulus reconstruction, it does not incorporate the distinct activation of different retinal cell types, and it does not provide a prediction of the effects of learning. A comparison or combination of the two approaches in future patient data may be valuable.

Image processing to optimize artificial vision. A potential enhancement of existing and future prosthesis technology is image processing prior to electrical stimulation, which could emphasize features of the visual scene of greatest significance to the patient, or more generally could produce a pattern of activation that is most likely to lead to accurate reconstruction. The rapid advance of cheap, low-power digital signal processing makes this kind of image processing much more accessible. Although simple ideas such as edge enhancement have obvious potential, there is currently no framework within which to systematically test their impact. In principle, the framework proposed here could be used to optimize pre-processing of the image. By evaluating patient performance on specific tasks (Fig. 6), and comparing to the predictions, the impact of pre-processing could then be assessed and fine-tuned.

Temporal responses and reconstruction. The present study tackled only the case of static images, for simplicity. A more complete understanding would require analyzing dynamic vision, with a

changing visual environment and head and eye movements. The methods presented translate naturally to the dynamic case. The simulation of RGC responses in the ISETBio software encompasses their dynamics, to the degree that current models can (Pillow et al 2008; Heitman et al. 2016). Based on dynamic responses, linear reconstruction of the input is performed by associating with each spike a spatiotemporal reconstruction filter, and summing those filters over space and time. As in the static case, estimation of the filters is obtained by least squares linear regression (Warland et al. 1997). In combination, these properties can be used to produce reconstructed movies that can then be compared to the original dynamic visual stimulus (see Results). For future work, some features of the dynamic simulation would need further investigation. First, although nonlinear reconstruction (see above) is readily accomplished with static images, this has not yet been performed with dynamic scenes, and the dimensionality of the problem is challenging. Second, perceptual measures of reconstruction accuracy, to improve upon the squared error metric, are less well developed for dynamic stimuli. Finally, appropriate stimuli are difficult to obtain, because the dynamic properties of vision due to eye movements are usually not present in the movie data sets that are readily available.

Interpreting the visual experience of patients. A challenge associated with assessing prosthesis function is attempting to understand what patients see. Verbal reports are helpful, but it is extremely difficult for the report provided by an untrained blind patient to give a complete sense of their visual experience. Because the reconstruction framework produces images as predictions of visual experience, it could be used by clinicians to compare with and interpret patient verbal reports. Such feedback may help to fine-tune the assumptions and parameters of the simulation (see Beyeler et al. 2017), such as the degree of ON dominance or learning (see above). Note that variations in performance that are relevant to the clinic, such as surgical details and missing ganglion cells due to degeneration, were not considered here.

Assessing the potential of technological developments. The development of retinal prostheses is ongoing, and represents a major international effort (Goetz and Palanker 2016). Given the enormous resources required to develop new technologies and obtain regulatory approval, as well as the risks to human patients receiving the implants, simulations that can predict reasonably accurately the performance of envisioned future devices could be immensely valuable (Beyleyer et al. 2017). The potential for such predictions is illustrated by examining the dependence of reconstruction on features such as electrode spacing and current spread (Fig. 5), which suggests that the biological limitations of the prosthesis need to be considered alongside the technical constraints.

Caveats and future. Several other uncertain features of the approach deserve mention. First, the concept underlying the approach – that the visual system reconstructs the visual image from the

retinal signal – is difficult to test explicitly. Second, the model of retinal activation by the photovoltaic implant and the model of healthy retinal processing are both rudimentary, and will need to be improved to better capture temporal phenomena, nonlinearities, noise, and cell type specificity. Furthermore, it will be challenging to produce an accurate simulation of the retinal activity produced by a specific prosthesis technology implanted in a specific patient with a particular disease process and treatment history. Finally, technical features of the reconstruction approach can surely be improved, as discussed above, and such improvements should be guided by patient reports over time. Thus, the framework presented here is not intended to be used unmodified in a clinical application. Instead, it may be considered a proof of concept and starting point for future efforts.

References

Angueyra JM, Rieke F. Origin and effect of phototransduction noise in primate cone photoreceptors. *Nat neurosci* 16:1692-700, 2013.

Beyeler M, Boynton GM, Fine I, Rokem A. pulse2percept: A Python-based simulation framework for bionic vision. *bioRxiv* 1:148015, 2017.

Boinagrov D, Lei X, Goetz G, Kamins TI, Mathieson K, Galambos L, Harris JS, Palanker D. Photovoltaic pixels for neural stimulation: circuit models and performance. *IEEE Trans. Biomed. Circuits Syst* 10:85-97, 2016.

Boinagrov D, Pangratz-Fuehrer S, Goetz G, Palanker D. Selectivity of direct and network-mediated stimulation of the retinal ganglion cells with epi-, sub-and intraretinal electrodes. *J Neural Eng* 11(2:026008, 2014.

Brainard DH, Jiang H, Cottaris NP, Rieke F, Chichilnisky EJ, Farrell JE, Wandell BA. Isetbio: Computational tools for modeling early human vision. [Abstract] Imaging Systems and Applications. Optical Society of America. IT4A-4, 2015. [The software is available at https://github.com/isetbio/isetbio]

Chambers JM. *Computational methods for data analysis*. New York: Wiley, 1977.

Chen SC, Suaning GJ, Morley JW, Lovell NH. Simulating prosthetic vision: I. Visual models of phosphenes. *Vision Res* 49:1493-506, 2009.

Chichilnisky EJ. A simple white noise analysis of neuronal light responses. *Network Comp Neural* 12:199-213, 2001.

Chichilnisky EJ, Kalmar RS. Functional asymmetries in ON and OFF ganglion cells of primate retina. *J Neuro* 22:2737-47, 2002.

CIE. International Commission on Illumination. *Fundamental chromaticity diagram with physiological axes – Part 1. Technical Report 170-1.* Vienna: Central Bureau of the Commission Internationale de l' Éclairage, 2006.

Cottaris NP, Elfar SD. How the retinal network reacts to epiretinal stimulation to form the prosthetic visual input to the cortex. *J Neural Eng* 2:S74, 2005.

Croner LJ, Kaplan E. Receptive fields of P and M ganglion cells across the primate retina. *Vision Res* 35:7-24, 1995.

Da Cruz L, Coley BF, Dorn J, Merlini F, Filley E, Christopher P, Chen FK, Wuyyuru V, Sahel J, Stanga P, Humayun M. The Argus II epiretinal prosthesis system allows letter and word reading and long-term function in patients with profound vision loss. *Br J Ophthalmol* 97:632-6, 2013.

Dacey DM. Primate retina: cell types, circuits and color opponency. *Prog Retin Eye Res* 18:737-63, 1999.

Dacey DM. Parallel pathways for spectral coding in primate retina. *Annu Rev Neurosci* 23:743-75, 2000.

Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. [Abstract] In Computer Vision and Pattern Recognition, IEEE Conference, 2009 P248-255, P 2009. [Software available online at http://image-net.org/download.php]

Eisen MD. Djourno, Eyries, and the first implanted electrical neural stimulator to restore hearing. *Otol Neurotol* 24:500-6, 2003.

Field GD, Chichilnisky EJ. Information processing in the primate retina: circuitry and coding. *Annu Rev Neurosci* 30:1-30, 2007.

Field GD, Gauthier JL, Sher A, Greschner M, Machado T, Jepson LH, Shlens J, Gunning DE, Mathieson K, Dabrowski W, Paninski L. Functional connectivity in the retina at the resolution of photoreceptors. *Nature* 467:673, 2010.

Goetz G, Smith R, Lei X, Galambos L, Kamins T, Mathieson K, Sher A, Palanker D. Contrast Sensitivity With a Subretinal Prosthesis and Implications for Efficient Delivery of Visual Information With a Subretinal Prosthesis. *Invest Ophthalmol Vis Sci* 56:7186-94, 2015.

Goetz GA, Palanker DV. Electronic approaches to restoration of sight. *Rep Prog Phys* 79:096701, 2016.

Gollisch T, Meister M. Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron* 65:150-64, 2010.

Hallberg LR, Ringdahl A. Living with cochlear implants: experiences of 17 adult patients in Sweden. *Int J Audiol* 43:115-21, 2004.

Heitman A, Brackbill N, Greschner M, Sher A, Litke AM, Chichilnisky EJ. Testing pseudo-linear models of responses to natural scenes in primate retina. *bioRxiv* 045336, 2016.

Humayun MS, Dorn JD, Da Cruz L, Dagnelie G, Sahel JA, Stanga PE, Cideciyan AV, Duncan JL, Eliott D, Filley E, Ho AC. Interim results from the international trial of Second Sight's visual prosthesis. *Ophthalmology* 119:779-88, 2012.

Jepson LH, Hottowy P, Mathieson K, Gunning DE, Dąbrowski W, Litke AM, Chichilnisky EJ. Focal electrical stimulation of major ganglion cell types in the primate retina for the design of visual prostheses. *J Neurosci* 33:7194-205, 2013.

Jepson LH, Hottowy P, Mathieson K, Gunning DE, Dąbrowski W, Litke AM, Chichilnisky EJ. Spatially patterned electrical stimulation to enhance resolution of retinal prostheses. *J Neurosci* 34:4871-81, 2014.

Jiang, H., Cottaris, N., Golden, J., Brainard, D., Farrell, J., & Wandell, B. Simulating retinal encoding: factors influencing Vernier acuity. *bioRxiv*, 109405, 2017.

Loudin JD, Simanovskii DM, Vijayraghavan K, Sramek CK, Butterwick AF, Huie P, McLean GY, Palanker DV. Optoelectronic retinal prosthesis: system design and performance. *J Neural Eng* 4:S72, 2007.

Lorach H, Goetz G, Smith R, Lei X, Mandel Y, Kamins T, Mathieson K, Huie P, Harris J, Sher A, Palanker D. Photovoltaic restoration of sight with high visual acuity. *Nat Med* 21:476-82, 2015a.

Lorach H, Goetz G, Mandel Y, Lei X, Kamins TI, Mathieson K, Huie P, Dalal R, Harris JS, Palanker D. Performance of photovoltaic arrays in-vivo and characteristics of prosthetic vision in animals with retinal degeneration. *Vision Res* 111:142-8, 2015b.

Mandel Y, Goetz G, Lavinsky D, Huie P, Mathieson K, Wang L, Kamins T, Manivanh R, Harris J, Palanker D. Cortical responses elicited by photovoltaic subretinal prostheses exhibit similarities to visually evoked potentials. *Nat Comm* 4:1980, 2013.

Mathieson K, Loudin J, Goetz G, Huie P, Wang L, Kamins TI, Galambos L, Smith R, Harris JS, Sher A, Palanker D. Photovoltaic retinal prosthesis with high pixel density. *Nat Photon* 6:391-7, 2012.

Menke W. *Geophysical data analysis: discrete inverse theory: MATLAB edition.* Waltham, MA: Academic press, 2012

Nirenberg S, Pandarinath C. Retinal prosthetic strategy with the capacity to restore normal vision. *PNAS* 109:15012-7, 2012.

Parthasarathy N, Batty E, Falcon W, Rutten T, Rajpal M, Chichilnisky EJ, Paninski L. Neural Networks for Efficient Bayesian Decoding of Natural Images from Retinal Neurons. *bioRxiv* 153759, 2017.

Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454:995, 2008.

Rieke F. *Spikes: exploring the neural code.* Cambridge: MIT press, 1999.

Shannon RV. Advances in auditory prostheses. *Curr Opin Neurol* 25:61, 2012.

Stingl K, Bartz-Schmidt KU, Besch D, Braun A, Bruckmann A, Gekeler F, Greppmaier U, Hipp S, Hörtdörfer G, Kernstock C, Koitschev A. Artificial vision with wirelessly powered subretinal electronic implant alpha-IMS. *Proc R Soc Lond B Biol Sci* 280: 20130077, 2013.

Stingl K, Bartz-Schmidt KU, Besch D, Chee CK, Cottriall CL, Gekeler F, Groppe M, Jackson TL, MacLaren RE, Koitschev A, Kusnyerik A. Subretinal visual implant alpha IMS–clinical trial interim report. *Vision Res* 111:149-60, 2015.

Waltzman SB, Cohen NL, Shapiro WH. The benefits of cochlear implantation in the geriatric population. *Otolaryngol Head Neck Surg* 108:329-33, 1993.

Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Sig Process* 13:600-12, 2004.

Wang L, Mathieson K, Kamins TI, Loudin JD, Galambos L, Goetz G, Sher A, Mandel Y, Huie P, Lavinsky D, Harris JS. Photovoltaic retinal prosthesis: implant fabrication and performance. *J Neural Eng* 9:046014, 2012.

Warland DK, Reinagel P, Meister M. Decoding visual information from a population of retinal ganglion cells. *J Neurophys* 78:2336-50, 1997.

Werginz P, Benav H, Zrenner E, Rattay F. Modeling the response of ON and OFF retinal bipolar cells during electric stimulation. *Vision Res* 111:170-81, 2015.

Zrenner E, Bartz-Schmidt KU, Benav H, Besch D, Bruckmann A, Gabel VP, Gekeler F, Greppmaier U, Harscher A, Kibbel S, Koch J. Subretinal electronic chips allow blind patients to read letters and combine them to words. *Proc R Soc Lond B Biol Sci* 278:1489-97, 2011.