Image Reconstruction from Cone Absorptions

Contents:

Introduction

Artificially generated vision has the potential to greatly improve quality of life for the 4.2 million Americans over the age of forty who suffer from legal blindness or low-vision [1]. The methods by which research groups and industries attempt to induce artificial percepts are widely varied in both stimulation (microelectrodes, photovoltaic, optogenetic) and in targeted layers (subretinal - bipolar cells, epiretinal - retinal ganglion cells) [2,3]. Nevertheless, it is widely understood that no matter the system of choice, the goal is to replicate the natural, physiological neural code of that layer as closely as possible. Given the diversity of cell types within the retina (~10 bipolar cell types [4] and ~20 types of retinal ganglion cell types [5]) and the inherent nonlinearities of cell behavior [6,7], replication of the neural code in response to infinite possible scenes is not at all straightforward. The matter is further complicated by the fact that multielectrode array geometry has yet been able to match the geometry of the retinal layers, meaning some cells are not accessible through stimulation, some electrodes can only stimulate two diametrically opposed cells simultaneously, or a portion of cells are impeded by non-targeted portions of the biology such as axon bundles which can stimulate axons of distal cells [8,9]. As a result, no existing prosthetic is able to perfectly match the native neural code in any retinal layer. Ideally, the resulting perception of imperfect stimulation would be simulated before the design and testing of these clinical implants in patients; however, this has largely not been the case. In this work, we propose a multiple linear regression model to simulate resulting perception via image reconstructions [10] from cone absorptions. We further investigate these image reconstructions by altering implant parameters such as position, size, distance, and stimulation selectivity to illustrate how simulation can inform designs of artificial vision implants.

Background

Scene to Cone Absorptions

In generating a transform from scene to cone absorptions or vice versa, we must summarize the main features of this system. In this case, a scene is an external image or 3D map that reflects light onto the eye. That scene is transformed by the optics of the eye into an optical image which projects onto the retina. The energy contained in that optical image is then absorbed by wavelength-dependent photoreceptors. Downstream retinal processes then perform cell-specific filtering and compression behaviors before the visual signal is sent into the cortex via action potentials along the optic nerve.

In the first stage, the scene is altered by various components of our eye depending on wavelength and luminance. For example, a small pupil diameter under bright conditions can result in greater diffraction which blurs the scene when imaged at our retina. Another common form of blurring in our vision is due to chromatic aberration, wherein the cornea and lens refract certain wavelengths more than others. For human optics, chromatic aberration results in medium wavelengths in greatest focus, followed by long wavelengths slightly out of focus, and short wavelengths even more out of focus. In addition to blurring, energy is also lost on the way to the retina in a wavelength-dependent manner. The lens and macula both absorb some of the light energy, disproportionately absorbing shorter wavelengths [11]. All of these different effects are captured by the complex-valued optical transfer function of the human eye modeled as a linear and shift-invariant system. The optical transfer function provides wavelength-dependent scaling and phase shifting across space. Note that the point-spread function is the spatial domain impulse response of the optical transfer function [11,12].

In the second stage leading to perception, photoreceptors transduce photons into chemical energy for processing by the rest of the retina. The two main types of photoreceptors are rods and cones, and the 3 subtypes of cones include long, medium, and short based on wavelength absorption efficiency [11]. For our purposes of considering restoration of useful vision in blind patients, cone photoreceptors are the most significant. High-acuity vision necessary for quality-of-life occurs under photopic (bright) conditions of 3 cd/m^2 or greater due to long and medium wavelength cones forming the majority of photoreceptors within the fovea [13]. The wavelength dependencies of the 3 cone types is also what grants us our color vision [11].

Following phototransduction by the cones and the production of their photocurrents, the visual information gets split into parallel visual streams where it is filtered and compressed through interneurons and retinal ganglion cells before being sent via action potentials to cell-specific portions of the cortex [11,14].

For our goal of image reconstruction in response to selectivity reduction, working at the level of the cones is most practical for a multitude of reasons. First, while filtering and compression of visual information downstream of the cones is useful for assessing contrast or providing adaptation, these manipulations also result in lossy compression which make cone responses the fundamental limit on our vision data according to information theory. This also implies that cones may be able to represent reduced selectivity in downstream neurons (bipolar cells and retinal ganglion cells); however, downstream neurons may not be able to represent reduced selectivity in the cones depending on the type of compression and filtering along the way [14,15]. Second, unlike many other cells in the retina, the cones behave approximately linearly under consistent photopic conditions with the number of photons transduced being proportional to the number of incident photons plus photon noise which follows the Poisson distribution. The SNR of the transduction increases with increasing brightness. Note that photocurrents do exhibit nonlinearities in amplitude with changing light levels, but the nonlinearity is static for each wavelength and cone type [11]. Since both the optical transfer function and cone phototransduction can be modeled as linear within our experimental domain, it is reasonable to apply multiple linear regression to learn models from cone absorptions to scenes for each combination of tested parameters. To note, in more complex conditions that apply dynamic changes in luminance, eye/head motion, and temporal frequency stimuli, a more involved model may be required past a linear regression.

Image Quality Metrics

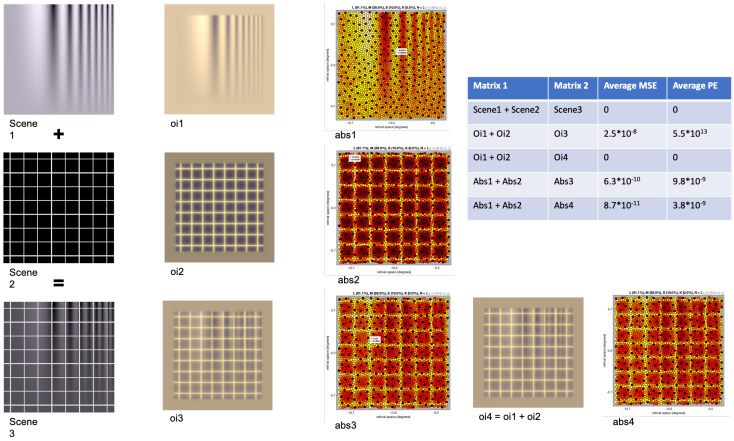
To assess image reconstruction quality, we chose two metrics that allowed us to train our multiple linear regression algorithm and assess perceptual image quality. The first metric used was mean squared error (MSE) which calculates the distance of all the points in an image reconstruction from the regression line. The multiple linear regression algorithm attempts to minimize mean squared error between the true images and the reconstruction. This method is a simple and common way to optimize the objective function of a linear model. Mean squared error alone is not normalized; therefore, it will depend on the number of pixels in the image being reconstructed. We present cone-to-pixel ratio and mean squared error per pixel in Table 1 to enable cross experiment comparisons. We also present the R^2 statistic which normalizes the mean square error by the degree of variance in the test data. While MSE and R^2 are common metrics for regressions, they are not the best metrics of image quality due to the structural nature of the images contributing to our perception of quality. To assess image quality with structural information maintained, we use the structural similarity index metric (SSIM). SSIM quantifies image degradation due to compression or processing by perceived change in structural information which incorporates important considerations of luminance and contrast rather than absolute error. SSIM is averaged between the ground truth images and the corresponding reconstructions and reported in Table 1.

Results

Linearity Verification

To verify that the system transform from scene to cone absorptions behaves linearly as predicted, we chose to perform a simple linearity check using the concept of superposition, wherein the net response of two or more stimuli is equivalent to the response to the sum of the stimuli. To start, we generated two individual scenes of the same size and over the same wavelengths. We chose the scenes ‘sweep frequency’ for the variation in spatial frequency (scene1) and the scene ‘gridlines’ for the sharp contrast (scene2) in order to stress alterations due to chromatic aberration or wavelength-dependent absorption. We then generated a third scene that was a summation of the first two using the function *sceneAdd.m* which combined the photon data of the two scenes (scene3). All scenes were set with a field-of-view (FOV) of 1.7° and a bit-depth of 32. As an initial sanity check, we verified that the mean-squared error (MSE) and percent error (PE) between scene3 and scene1 + scene2 was 0. Next, we pushed all three

scenes through human optics (generated via *oiCreate('human')*) with the FOV again set to 1.7°. As an additional comparison, we then added the optics (*oiAdd.m*) from the first scene (oi1) and the optics of the second scene (oi2) with weights equal to 1 to create a fourth optical image (oi4). The optical images oi3 and oi4 were then compared to oi1 + oi2. As expected, oi4 and oi1 + oi2 had a PE and MSE of 0. However, oi3 (scene3 pushed through human optics) and oi1 + oi2 had a reasonable PE of $2.4*10^{-8}$ but a MSE of $5.5*10^{13}$. While the MSE appears large at first, we must consider that MSE is not normalized. Because the magnitude of photons in the optical images is 10^{13}, the average MSE is reasonably small. Finally, we generated a cone mosaic that was 1.7°x1.7° FOV, -10° in x-direction of eccentricity (raphe region), and had an integration time of 10ms which is the estimated visual integration window for cones [16]. We computed cone absorptions from each of these optical images providing us abs1, abs2, abs3, and abs4. Comparing abs1 + abs2 to abs3, we found a PE average of $9.8*10^{-9}$ and MSE average of $6.3*10^{-10}$. Comparing abs1 + abs2 to abs4, we found a PE average of $3.8*10^{-9}$ and MSE average of $8.7*10^{-11}$. Plotted results can be seen in Figure 1. All these values suggest that a linear transformation should be able to adequately represent the multiple stages from scene to optical image to cone absorptions under our experimental conditions of constant display, luminance, and stable eye position.



Matrix 1	Matrix 2	Average MSE	Average PE
Scene1 + Scene2	Scene3	0	0
Oi1 + Oi2	Oi3	$2.5*10^{-8}$	$5.5*10^{13}$
Oi1 + Oi2	Oi4	0	0
Abs1 + Abs2	Abs3	$6.3*10^{-10}$	$9.8*10^{-9}$
Abs1 + Abs2	Abs4	$8.7*10^{-11}$	$3.8*10^{-9}$

Figure 1: Scenes, Optical Images, and Cone Absorptions for Linearity Verification. Error Metrics summarized in Inset Table.

Generating Data

To train the multiple regression model, we needed to source a large number of images and transform those images into cone absorptions. We chose ImageNet as our source of images due to their extensive library (14,197,122 possible images) and reliable quality. ImageNet provides only grayscale images which is preferred. Thus far, all artificial vision implants only reliably reproduce light and dark percepts [17]. From this extensive library, we selected 5500 images that were readily available to us. This selection was then reduced to 4096 images which could be cropped to the same dimensions (300x450). The dimension matching allows us to know that each image will have the same field of view for a given display and distance from the lens. This selection of images was randomized to remove the chance of structural similarity in an incrementing subset of the cropped image library. The library was then split 3:1 into a training and testing set. Ultimately, we used 2800 images for training and 1008 images for testing due to computational speed limitations.

In addition to selectivity, we chose a number of parameters to vary when considering image reconstruction. Namely, we varied the position (eccentricity in degrees) of the cone mosaic of interest from fovea to periphery ([0,0],[-5,0],[-10,0],[-15,0]). For reference, [-10°,0°] indicates the raphe region [18] of the retina which is the lowest eccentricity most epiretinal implants are able to insert. For each of these eccentricities, we maintained a distance from the display image at 0.25m and a cone mosaic size of 1.7°x1.7°. We also swept cone mosaic size ([0.5x0.5], [1x1], [1.7x1.7], [2x2], [3.4x3.4]). For reference, [3.4°x3.4°] is equivalent to a [1mmx1mm] implant assuming a 17mm retina diameter. However, on a minor note, for the other sweeps, we maintained size at 1.7°x1.7° due to the computational loads of many cones at larger sizes. For the size sweep, we maintained position at [-10°,0°] and distance from image at 0.25m. Finally, we also swept distances from the display to the lens (0.25m, 0.3m, 0.4m, and 0.6m), while maintaining a size of 1.7°x1.7° and an eccentricity of [-10°,0°]. For reference, a 0.6m distance from the image to the lens is a recommended distance from a computer screen [19] while a 0.25m distance allows a 13.6° horizontal fov for 450 pixels (although this distance would require stereoscopic convergence in practice).

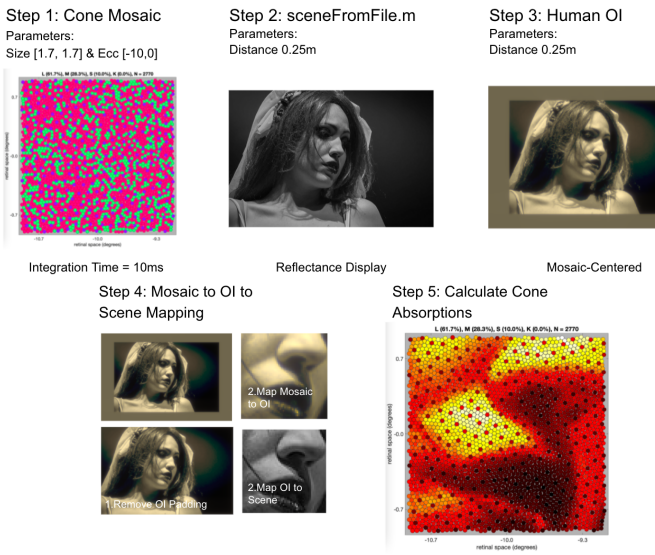


Figure 2: Pipeline for generating scene-absorption training and testing sets using ISETBio tools

For each set of parameters, we generated a cone mosaic with an integration time of 10ms (again, chosen due to the visual integration window [16]). We also generated a human optical image for the given distance parameter. Each image was loaded as a scene object using the *sceneFromFile.m* function at the specified distance parameter. The display was specified as the 'reflectance-display' which has RGB properties '*rendered under a D65 illuminant and the surface reflectances fall within a 3D linear model of natural surface reflectances*,' according to the ISET Cam [20] *sceneFromFile.m* description. Each of the 3808 images was loaded as a scene individually and pushed through the human optics computation (oiCompute.m). Then, two independent stages occurred. In one stage, the optical image was cropped to remove padding (*oiCrop.m*), allowing the optical image to have a 1:1 relationship with the original scene. From this, we could then look at the recorded square map of the optical image applied to the cone mosaic to determine the corresponding square map of the scene. This allowed us to isolate the portion of the scene incident on the cone mosaic. In a second stage, the optical image went through computations to generate cone absorptions. Each of the images was kept mosaic-centered for two reasons: 1) each cone mosaic receives the same training and test images with variations only due to size of the mosaic and 2) provides greater flexibility in eccentricities as -15° in the x-direction would fall outside the maximum horizontal fov of the scene on the display. This data generation pipeline is demonstrated in Figure 2. For each set of parameters, we have a single cone mosaic, 2800 image-absorption pairs for training, and 1008 image-absorption pairs for testing. Summary of the sweeps can be found in Figure 3.

Eccentricity (Deg)	[0,0]	**Mosaic Dimensions (Deg)**	0.5x0.5	**Distance (m)**	0.25
	[-5,0]		1.0x1.0		0.3
	[-10,0]		1.7x1.7		0.4
	[-15,0]		2.0x2.0		0.6
Constant Params: Size (Deg) & Dist (m)	1.7x1.7 & 0.25		3.4x3.4		
		Constant Params: Ecc (Deg) & Dist (m)	[-10,0] & 0.25	**Constant Params:** Size (Deg) & Ecc (Deg)	1.7x1.7 & [-10,0]

Figure 3: Data generation sweeps over implant position (left), implant size (center), and scene distance (right) holding all other parameters constant

Multiple Linear Regression

Once cone absorption data was generated across a variety of hyperparameters, we generated multiple linear regression models with a bias term to predict the original grayscale images displayed on the monitors from the cone absorption vectors. The number of cones in the window of the retina was a function of the eccentricity and the size of the window. This information is summarized in Table 1. The larger the implant size and the closer to the fovea, the more cones were present in the cone mosaic window.

As described previously, all images projected on the ISETBio simulated monitor were cropped to a size of 300x450. However, given the parameters of the experiment, only a subset of the pixels on the monitor have an impact on the cone mosaic of interest. Thus, it is incorrect to expect the linear regression model to accurately predict all 135,000 of the pixels displayed on the monitor. As such in the data generation step, only the sub-window of the grayscale image that correspond with the ISET optical image projected onto the mosaic is reported to the multiple linear regression dataset.

The multiple linear regression model seeks to predict the matrix of flattened grayscale images, Y, from the matrix of their corresponding cone absorptions, X. We sought a model of the form Y = A * X + b, choosing A and b which minimize the L2 norm between predicted Y and the ground truth images. Call X' and Y' the X and Y matrices adjusted so that all feature means were zero. Then we can find the A matrix and bias vector as follows:

$$A = (Y'X'^{T})*(X'X'^{T})^{-1}$$

$$b = mean(Y) - A*mean(X)$$

For each set of hyperparameters, the training data was used to generate a multiple linear regression model of the above form. To ensure the same sub-window of training images was used in all hyperparameter sets, the images were mosaic-centered on the display monitor in accordance with the implant eccentricity under inspection. This way, accuracy comparisons can be drawn across different experiments with respect to a single image. The following sections present the results associated with sweeping a single hyperparameter, and all the data is summarized in Table 1.

Monitor Distance (m)	Square Window Dimension (degrees)	Horizontal Eccentricity (degrees)	Train Data MSE per pixel	Test Data MSE per pixel	Train Data SSIM	Test Data SSIM	Train Data R^2	Test Data R^2	Number of Cones	Square Image Dimension (Pixels)	Cone to Pixel Ratio
0.25	1.7	0	376.35	861.335	0.5978	0.4189	0.9141	0.8121	25268	56	8.0574
0.25	1.7	-5	410.339	886.7892	0.5776	0.4059	0.9063	0.8065	4807	56	1.5328
0.25	1.7	-10	409.6634	926.8383	0.5744	0.3883	0.9062	0.7969	2770	56	0.8833
0.25	1.7	-15	440.09	913.66	0.5615	0.3944	0.8996	0.8006	2185	56	0.6967
0.25	0.5	-10	509.9682	655.8379	0.5694	0.4996	0.8855	0.8496	249	17	0.8616
0.25	1	-10	501.743	710.6643	0.5705	0.4891	0.8856	0.8444	976	33	0.8962
0.25	2	-10	399.4043	999.2089	0.5717	0.3639	0.9089	0.782	3851	66	0.8841
0.25	3.4	-10	259.5	1353.1	0.6212	0.2559	0.941	0.7051	11077	112	0.8831
0.3	1.7	-10	428.6926	900.1621	0.5677	0.3989	0.9022	0.8036	2770	56	0.8833
0.4	1.7	-10	428.7466	899.0698	0.5679	0.399	0.9021	0.8038	2770	56	0.8833
0.6	1.7	-10	428.8976	898.0745	0.5679	0.3991	0.9021	0.804	2770	56	0.8833

Table 1: Summary of Multiple Linear Regression Model Results

Implant Eccentricity

In this section, we characterize the accuracy of the linear model in various regions of the retina, including the fovea [0, 0], raphe [-10, 0], and periphery[-15, 0]. We expect that the reconstruction accuracy should be correlated with the cone-to-pixel ratio where increases in this ratio correspond with better reconstructions. Therefore, we expect better reconstruction assessments towards the fovea over the periphery. This is the exact relationship which we see demonstrated in Figure 4 in both the train and test data. Figure 5 demonstrates the number of cones in a square mosaic of visual angle dimension 1.7°x1.7° in each of these regions. As expected, cone density is highest in the fovea and drops off as you get closer to the periphery. Because the set of pixels consequential to image reconstruction is 56x56 in all cases, the cone-to-pixel ratio also increases in the direction of the fovea.

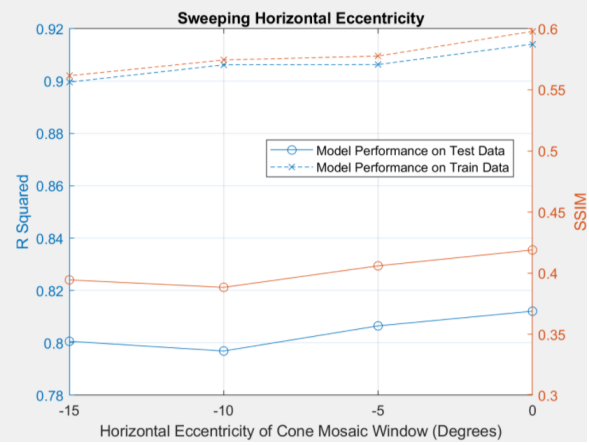


Figure 4: Impact of implant eccentricity on model fit (R-squared) and quality of image reconstruction (SSIM)

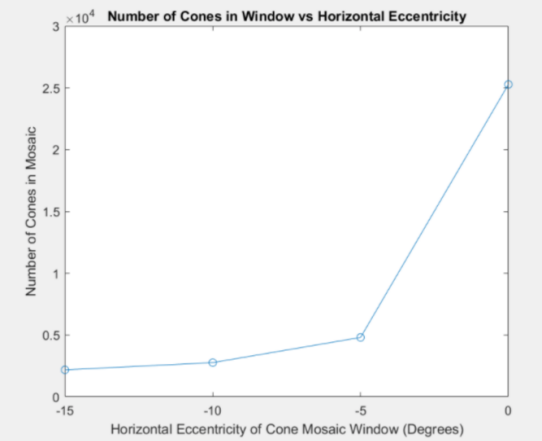


Figure 5: Relationship between implant eccentricity and respective cone density

Figure 6 shows an example of the reconstruction on three chosen test images in each of the eccentricities swept. The image quality reconstruction reduces slightly from left (foveal) to right (peripheral). This suggests that the rank of the covariance matrix of the image data we used is low enough to be captured by the cone density in the raphe and periphery, such that an increase in the number of the cones only marginally improves the performance.

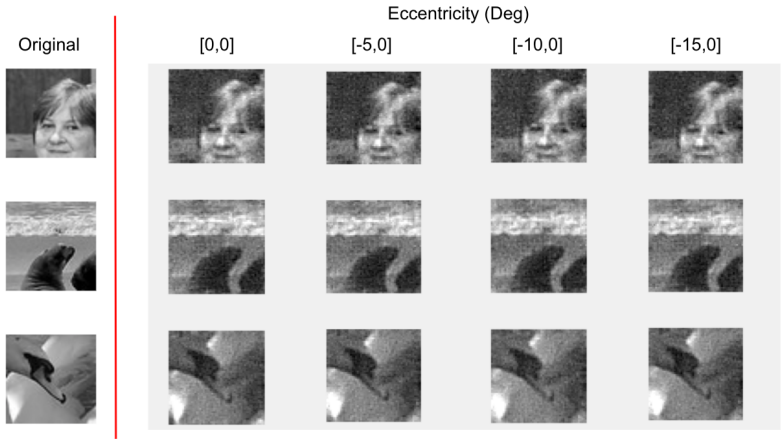


Figure 6: Image reconstruction examples for models generated at different eccentricities

Implant Size

In this section, we keep the monitor 25cm away from the retina and the mosaic of interest centered in the raphe, eccentricity [-10°, 0°] and vary the size of the mosaic from 0.5° to 3.4°. The larger the size of the mosaic, the larger the number of cones and the larger the image of consequence to those cone absorptions. The image size increasing is reflected in Figure 9. As the window size gets bigger, so do the number of pixels from the image on the monitor, reflecting the need for noramlized metrics for comparison as discussed earlier.

As demonstrated in Figure 8, the cone density and as follows the cone-to-pixel ratio is roughly constant over the sweep, with a slight downwards trend as window size increases. This is consistent with the accuracy of the linear models trained on the test data as seen in Figure 7. However, the accuracy of the linear model on the train data seems to proceed in the opposite direction as on the training data. We suspect the performance trend on the training data is attributed to over fitting, since larger window sizes provide more weights that can be trained in the matrix.

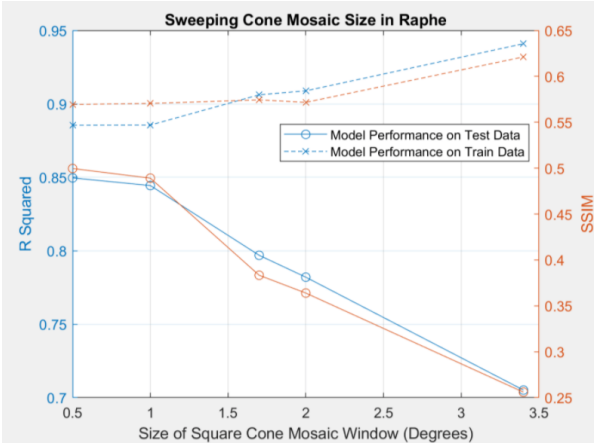


Figure 7: Impact of implant size on model fit (R-squared) and quality of image reconstruction (SSIM)

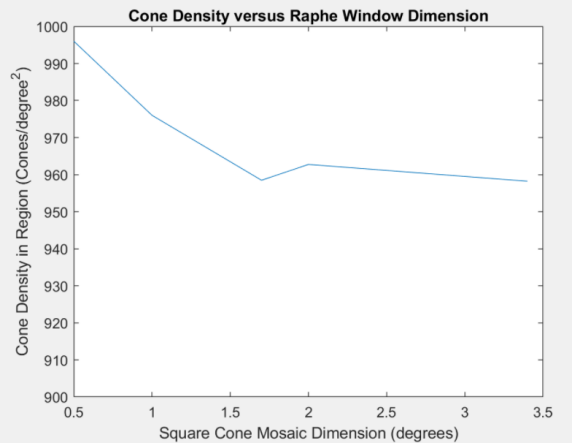


Figure 8: Relationship between implant size and respective cone density

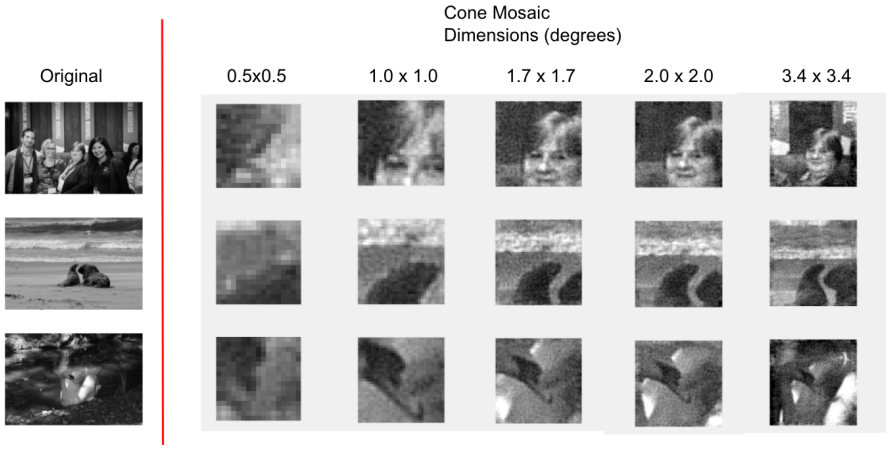


Figure 9: Image reconstruction examples for models generated at different implant sizes

Scene Distance

In this section we sweep the monitor distance from 25cm to 60cm while keeping the cone mosaic centered in the raphe ([-10°, 0°]) with a constant dimension of 1.7°x1.7°. We expect as distance is increased, the cone-to-pixel ratio would decrease, and the resulting reconstructions would be reduced in quality. However, the results for this sweep are inconclusive. There is no trend demonstrated in Figure 10, and as Table 1 demonstrates, even with adjusting the the distance in both the scene generation and optical image inputs of ISETBio, the image of interest for each of the monitor distances remained 56x56. However as the monitor gets further away, the size of the image of interest should increase. More investigation is required to validate the results of this experiment.

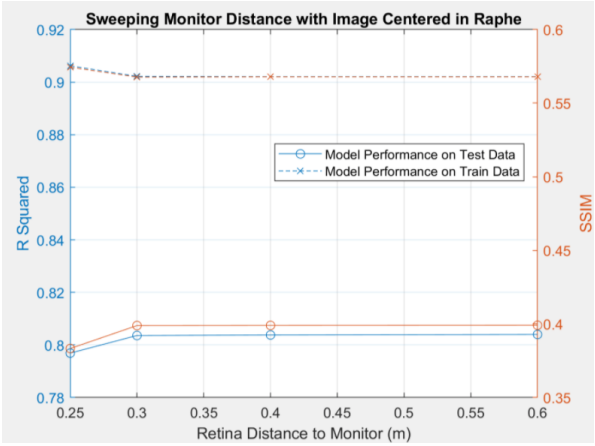


Figure 10: Impact of scene distance on model fit (R-squared) and quality of image reconstruction (SSIM)

Model Evaluation

In Figure 11, we have plotted the structural similarity between the original image and its reconstruction for the test images for the linear model experiment for the raphe at 1.7°x1.7° with the monitor 25cm away. We plot this structural similarity against the mean pixel value of the test images as a proxy for brightness. The trend in the scatter plot clearly show the best performance is for medium light level images. This can be attributed to the bounds of the mean value being at 0 and 256. We found that for all linear models trained, the computed bias term was a near constant vector of 45, such that no cone absorption would lead our linear models to predict the original image was a dark shade of gray. This goes sharply against intuition that cone absorptions should be a linear function of the original function without a bias term and warrants further investigation. Adjustment of scene inputs to include a darkest pixel value black square and a brightest pixel value white square may provide the final tuning necessary to optimize reconstruction in these extreme regimes.

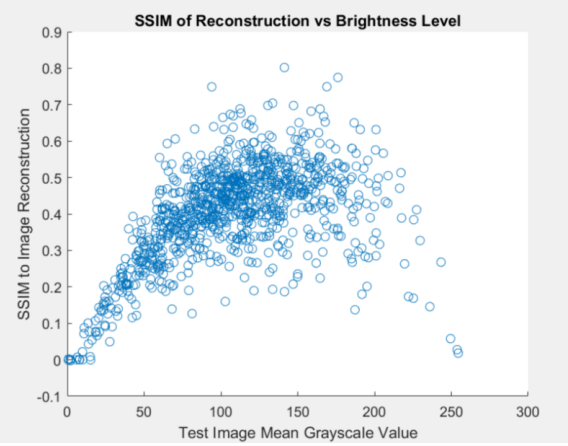


Figure 11: Evaluating model performance as a function of image statistics

Neural Networks

In parallel with training the multiple linear regression model, we also conducted a search for models capable of handling nonlinearities which would offer improved model fitting as we increase the complexity of our investigations. This was primarily motivated by the expected development of ISETBio tools capable of extending visual representations to the retinal ganglion cell layer, as well as the potential incorporation of other factors such as stimuli of varying temporal frequencies, eye movements, and dynamic changes in luminance level. Figure 12 demonstrates several approaches considered in order to represent this transformation in a format conducive to training neural networks.

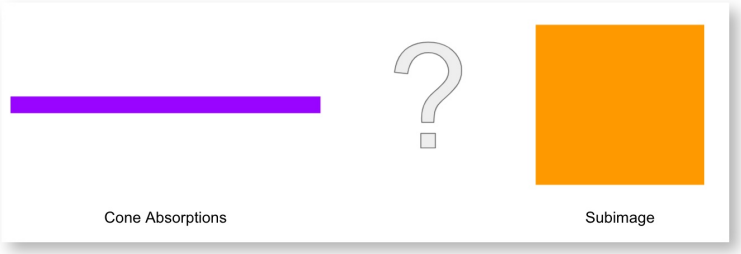


Figure 12a

Figure 12b

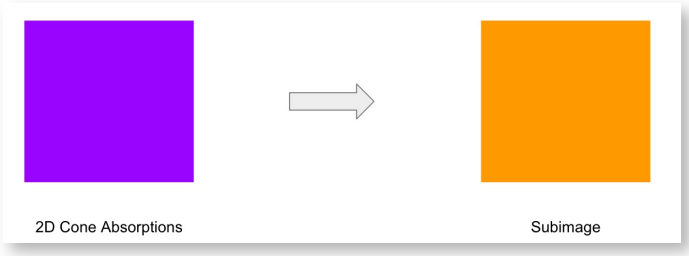


Figure 12c

Figure 12: Alternative modeling approaches for neural network design

As shown in Figure 12a, our learning problem can be abstracted to an input vector of features and a 2D output matrix. One can imagine a simplistic approach whereby our 2D output matrix is flattened down to one dimension (Figure 12b), facilitating the use of a multilayer perceptron (MLP) architecture with a regression output layer. Figure 13 reveals the base MLP architecture considered for fitting our data consisting of a single hidden layer. Many variations of this MLP architecture were trained, including the addition of more hidden layers, swapping out nonlinear activation functions, and a range of training hyperparameters such as learning rate and schedule. These attempts ultimately produced lackluster results, leading us to an alternative formulation of the learning problem that incorporates more of the useful information in the data.

ANALYSIS RESULT

	Name	Type	Activations	Learnables
1	input 4807 features	Feature Input	4807	-
2	fc 3136 fully connected layer	Fully Connected	3136	Weights 3136×4807 Bias 3136×1
3	batchnorm Batch normalization	Batch Normalization	3136	Offset 3136×1 Scale 3136×1
4	relu ReLU	ReLU	3136	-
5	regressionoutput mean-squared-error	Regression Output	3136	-

Figure 13: Base MLP architecture used for fitting nonlinear model

As shown in Figure 12c, one can imagine a slightly more complex approach whereby our input vector of features is reshaped into a 2D input matrix. After all, there is spatial information between the cone absorptions, and taking advantage of such structure may lead to a more successful model. Figure 14 reveals the more complicated base convolutional neural network (CNN) architecture with a regression output layer. Once again, we experimented with many variations of the CNN architecture including the number of layers, the size and number of convolution kernels, and the training hyperparameters. These attempts also generated poor-fitting models. Sticking with this CNN formulation, we also considered leveraging pre-trained networks such as those that perform semantic segmentation to conduct transfer learning, but ran into a series of problems with practical formulation of the data in a manner that could take advantage of the learning already conducted within the pre-trained networks.

ANALYSIS RESULT

	Name	Type	Activations	Learnables
1	imageinput 52×52×1 images with 'zerocenter' normalization	Image Input	52×52×1	-
2	conv_1 8 5×5 convolutions with stride [1 1] and padding 'same'	Convolution	52×52×8	Weights 5×5×1×8 Bias 1×1×8
3	batchnorm_1 Batch normalization	Batch Normalization	52×52×8	Offset 1×1×8 Scale 1×1×8
4	relu_1 ReLU	ReLU	52×52×8	-
5	conv_2 16 5×5 convolutions with stride [1 1] and padding 'same'	Convolution	52×52×16	Weights 5×5×8×16 Bias 1×1×16
6	batchnorm_2 Batch normalization	Batch Normalization	52×52×16	Offset 1×1×16 Scale 1×1×16
7	relu_2 ReLU	ReLU	52×52×16	-
8	conv_3 32 5×5 convolutions with stride [1 1] and padding 'same'	Convolution	52×52×32	Weights 5×5×16×32 Bias 1×1×32
9	batchnorm_3 Batch normalization	Batch Normalization	52×52×32	Offset 1×1×32 Scale 1×1×32
10	relu_3 ReLU	ReLU	52×52×32	-
11	conv_4 1 5×5 convolutions with stride [1 1] and padding 'same'	Convolution	52×52×1	Weights 5×5×32 Bias 1×1
12	batchnorm_4 Batch normalization	Batch Normalization	52×52×1	Offset 1×1 Scale 1×1
13	relu_4 ReLU	ReLU	52×52×1	-
14	dropout 20% dropout	Dropout	52×52×1	-
15	regressionoutput mean-squared-error	Regression Output	52×52×1	-

Figure 14: Base CNN architecture used for fitting nonlinear model

The failure of these efforts to produce competitive models revealed several useful insights about the nature of the data utilized for this project. First, learning a model from cone absorptions to subimages is a fundamentally large undertaking considering the dimensionality of both the cone absorptions and subimage pixels are on the order of thousands. This dimensionality compounded with the relatively small number of training instances in our training data set may have placed the neural network approach with its much larger expressivity at a distinct disadvantage. Second, this result caused us to conduct a more rigorous assessment of the linearity of the transforms from scene to cone absorptions in ISETBIO. Given our verification above of the linearity of the the transforms from scene to cone absorptions in ISETBIO, we decided that, while pursuit of nonlinear models may become necessary in the future to handle more complex simulations, we could take advantage of the present multiple regression to explore how we might utilize such models to investigate implant stimulation selectivity.

Simulating RGC Mappings

As motivated in the Introduction section, the mismatch between the neural circuitry at the retinal ganglion cell layer and the geometry of an electrode array means that we must engineer solutions with imperfect selectivity of RGCs. For example, an axon bundle may prevent unique electrical access to a target cell without simultaneous activation of distal axons from many other RGCs. Our inability to evoke exact neural codes necessarily impacts the eventual percept generated by the visual cortex. It would be highly desirable to simulate situations involving imperfect selectivity of RGCs and assess the resulting image reconstruction predicted by an accurate model of perception. These results could help inform our engineering approaches by determining the most and least likely selectivity situations for improving visual perception.

The process that we used to associate cone absorptions with retinal ganglion cells is presented in Figure 15. During Step 1, we access the cone mosaics corresponding to experiment parameters as described in the Data Generation section above. In Step 2, we take advantage of existing RGC code in ISETBIO to generate midget RGC mosaics spanning the same spatial extent as the cone mosaic in the previous step. We perturb the midget RGC mosaic generation slightly so as to generate a second distinct mosaic, corresponding to hypothetical mosaics for both ON and OFF midget cell types. Step 3 uses region of interest code from ISETBIO to map the receptive fields of the midget RGCs (the circles tiling the retinal space in Step 2) to the underlying sets of cones. Step 3 of Figure 15 visually demonstrates a 95% selectivity rate for both the ON and OFF midget mosaics. In this example, the cones highlighted cyan and magenta were chosen to undergo cone absorption modifications as the non-selective regions of the ON and OFF midget RGC mosaics, respectively. Note that for these experiments, we assigned non-selectivity randomly to cells within the mosaics, and that the absorptions for the corresponding cones were zeroed to represent the inability of stimulating electrodes in activating the RGCs. Of course, with more accurate RGC modeling, one could more diligently assign non-selectivity regions according to geometries commonly observed in *ex vivo* experimentation.

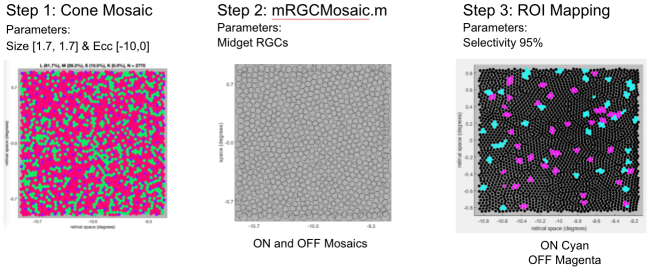


Figure 15: Pipeline for modifying cone absorptions by mapping RGC mosaics to cone mosaics using ISETBIO tools

Figure 16 denotes the experiments conducted under these modified cone absorption conditions. It is important to note that we still use the same models trained by multiple regression as discussed above, but now we modify cone absorptions in the testing data set according to the provided RGC selectivities and we assess the resulting image reconstructions. Given our reasonably accurate models of perception

using cone absorptions, the idea is that modifying these cone absorptions and then applying the same model of perception may approximate how the visual cortex responds to novel RGC firing activity resulting from imperfect stimulation. The sweep over RGC selectivity looks to characterize how progressive reductions in selectivity change the image reconstruction of the multiple regression model with default parameters (0.25m distance from scene, 1.7°x1.7° implant size, [-10°,0°] implant eccentricity). The sweep over eccentricity investigates whether a constant lack of selectivity changes the results of the eccentricity analysis performed above under total selectivity.

RGC Selectivity (%)	100
	99
	95
	90
	80
Constant Params: Size (Deg) & Dist (m) & Ecc (Deg)	1.7x1.7 & 0.25 & [-10,0]

Eccentricity (Deg)	[0,0]
	[-5,0]
	[-10,0]
	[-15,0]
Constant Params: Size (Deg) & Dist (m) & Sel (%)	1.7x1.7 & 0.25 & 95

Figure 16: Testing set sweeps over RGC selectivity (left) and implant eccentricity under reduced RGC selectivity (right) holding all other parameters constant

Implant Eccentricity

This experiment analyzes how quality of image reconstruction varies with different implant eccentricities, this time after modifying the cone absorptions in each testing set to represent a constant 95% RGC selectivity. Figure 17 demonstrates that, unsurprisingly, the quality of image reconstruction as measured by the SSIM metric is universally worse than the image reconstructions generated by the same model with unmodified cone absorptions representing full RGC selectivity. However, one very interesting trend that emerges from the plot is an increased sensitivity to implant eccentricity. In particular, as we move out from the fovea toward the periphery, the absolute quality of image reconstruction drops more rapidly than seen in the previous experiment. While image quality may not suffer greatly as we move further out in eccentricity across the retina under assumptions of perfect selectivity, in the more realistic realm of non-selective stimulation we find that eccentricity appears to become a more important factor. It is this type of result that could be leveraged in the future to guide decisions about implant parameters during the design process. Figure 18 once again shows our three cropped original scenes from the testing data set and provides a visualization of our resulting image reconstructions across eccentricities. The degradation in image quality from 0 degrees toward -15 degrees is visually significant.

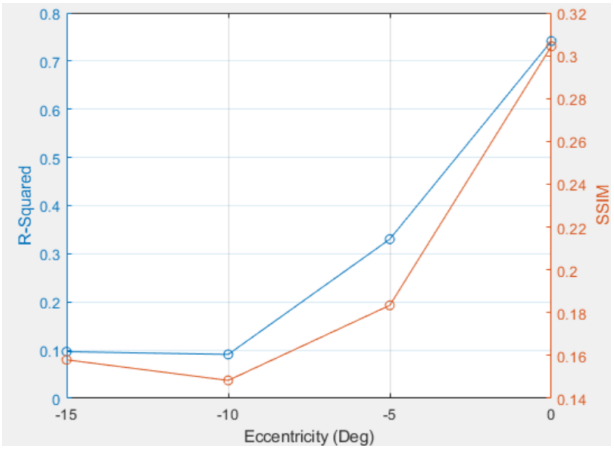


Figure 17: Testing set model fit (R-squared) and quality of image reconstruction (SSIM) at different eccentricities after modifying cone absorptions under 95% RGC selectivity

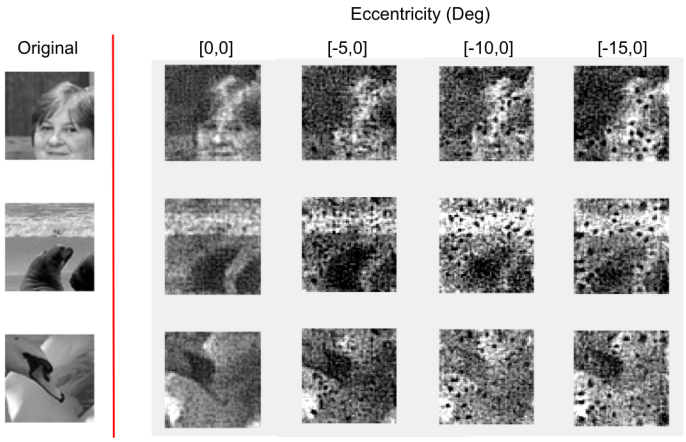


Figure 18: Image reconstruction examples for models generated at different eccentricities after modifying cone absorptions under 95% RGC selectivity

Implant RGC Selectivity

This experiment analyzes how quality of image reconstruction changes with decreasing RGC selectivity by modifying the cone absorptions of our default parameter testing set to varying degrees. Figure 19 demonstrates a very steep decrease in the SSIM metric during the initial decrease from 100% selectivity down to 95% and 90%, then levels off at a very poor SSIM value under 0.1 as the RGC selectivity drops under 80%. This decreasing trend is intuitive, as any initial perturbation to an image is likely to present a more significant perceptual change than once the image is already significantly degraded. The trend is confirmed using the visualizations provided in Figure 20, whereby we observe that decreasing the RGC selectivity ultimately renders the image reconstruction unrecognizable. This is also intuitive, as there is bound to be some selectivity threshold below which percepts cannot be reliably evoked due to lack of RGC activation control. It is important to note that our RGC to cone mapping is only a first-order approximation for RGC selectivity, so one should take the absolute results of these investigations with a grain of salt. Nonetheless, the trends evoked here do suggest that this method of simulation can give us useful insights into how RGC selectivity may qualitatively and quantitatively affect the reconstruction of an image.

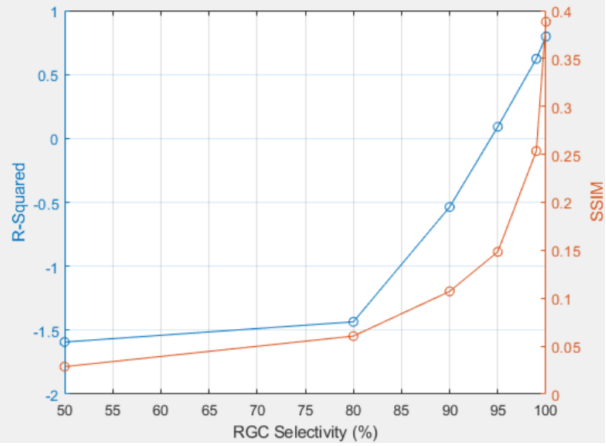


Figure 19: Testing set model fit (R-squared) and quality of image reconstruction (SSIM) for default model parameters at different RGC selectivities after modifying cone absorptions

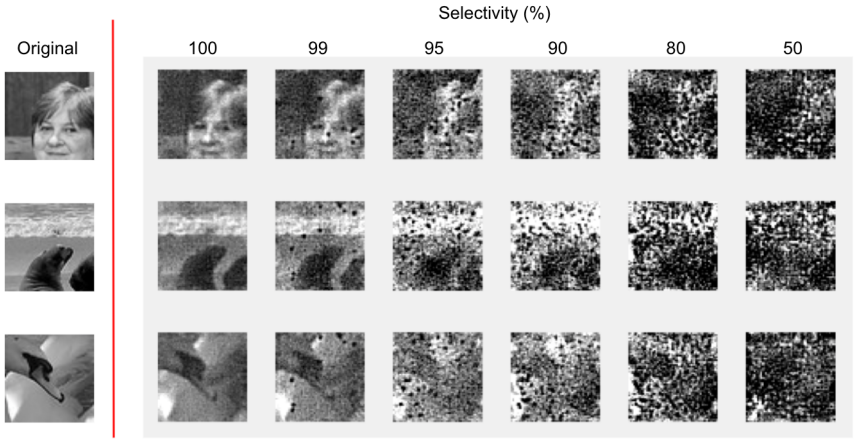


Figure 20: Image reconstruction examples for default model parameters at different RGC selectivities after modifying cone absorptions

Conclusions

From our testing of linearity, we can conclude that under our simulated conditions, a linear regression model is all that is needed to adequately predict an image from cone absorptions (taking advantage of the shift-invariant optical transfer function across wavelengths). This greatly reduces computational loads and provides greater abstraction to biology.

The performance of the linear models we trained are related to the cone density, or cone-to-pixel ratio, of the mosaics utilized in each of the experiments. Increasing cone density leads to better reconstruction performance. The presence of a non-zero bias term in the multiple linear regression and unchanging image size with changing distance are causes for further investigation.

Although the neural networks that we trained performed relatively poorly as models for our data, the investigation was a healthy exercise in modeling the problem and may become useful again as we increase the complexity of the simulation environment.

Our investigation of RGC selectivity demonstrated the capability of simulation models to provide actionable insights that could guide engineering decisions relevant to a future implant. We found that parameters such as eccentricity may become important factors in image reconstruction performance when coupled with reduced selectivity. We also found that sweeping selectivity can yield useful assessments of how image reconstruction performance may change with differing RGC selectivity capabilities.

It is important to acknowledge the limitations of the work presented above. Such limitations are closely intertwined with the Future Directions section below and provide a context with which to interpret our results.

Limitations

First, our choice to use images for data generation necessitated the selection of a display on which to present each scene in the ISETBIO environment. In order to best mimic the natural scenes in our dataset, we opted for the 'reflectance display' to use CIE standard illuminant D65. Second, our models were trained from cone absorptions only without consideration for rods. While this simulation omission may be reasonably justified for the photopic domain used throughout this project, it would become an important gap in our simulation capabilities when presenting scotopic scenes. Third, the RGC mosaics generated in the latter portion of our project could only be modeled after midget retinal ganglion cells, which are certainly necessary but not sufficient to capture the multitude of cell types, morphologies, and properties within the RGC layer of the retina. We can expect the usefulness of insights derived from investigations like the ones presented here to increase as we incrementally address these limitations of our simulation pipeline.

Future Directions

We hope to continue developing the methods discussed in this project into an increasingly functional simulation environment. Primarily, we look forward to adapting our codebase from cone simulations in the photoreceptor layer of the retina to retinal ganglion cell simulations as pending ISETBIO tools become available to us. We hope that this improvement will help us introduce more realistic selectivity questions to help guide research efforts. For example, which of these selectivity challenges is expected to be more perceptually significant: avoiding the activation of axon bundles or isolating proximal ON and OFF cell types? Another future direction to help improve our understanding of how an implant might facilitate perception across time is the introduction of eye movements and microsaccades to our data generation pipeline.

Acknowledgements

Thanks to Dr. Brian Wandell, Dr. Joyce Farrell, and David Cardinal for sharing expertise throughout the duration of this project. Thanks to Nicolas Cottaris for his work on the retinal ganglion cell and region of interest functions in ISETBIO. Thanks to Nicolas Cottaris for his work on the retinal ganglion cell and region of interest functions in ISET Bio.

References

ISET Bio:

https://github.com/isetbio/isetbio/wiki

Nicolas P. Cottaris, Haomiao Jiang, Xiaomao Ding, Brian A. Wandell, David H. Brainard; A computational-observer model of spatial contrast sensitivity: Effects of wave-front-based optics, cone-mosaic structure, and inference engine. Journal of Vision 2019;19(4):8. doi: 10.1167/19.4.8.

[1] CDC. *Common Eye Disorders and Diseases*. Source: Division of Diabetes Translation, National Center for Chronic Disease Prevention and Health Promotion. 03 June 2020. https://www.cdc.gov/visionhealth/basics/ced/index.html

[2] Chen, Z.C., Wang, B-Y, & Palanker, D. Real-Time Optimization of the Current Steering for Visual Prosthesis. *Biological Physics.* (2020)

arXiv:2012.13467

[3] Chuang, AT, Margo, CE, and Greenberg, PB. Retinal implants: a systematic review. B J Ophthalmol. 2013; 28(7):852-856.

[4] Yan, W., Peng, YR., van Zyl, T. *et al.* Cell Atlas of The Human Fovea and Peripheral Retina.*Sci Rep* 10, 9802 (2020). https://doi.org/10.1038/s41598-020-66092-9

[5] Kim, U. S., Mahroo, O. A., Mollon, J. D., & Yu-Wai-Man, P. Retinal Ganglion Cells-Diversity of Cell Types and Clinical Relevance. *Frontiers in neurology*, *12*, 661938. (2021) https://doi.org/10.3389/fneur.2021.661938

[6] Helene Marianne Schreyer, Tim Gollisch,Nonlinear spatial integration in retinal bipolar cells shapes the encoding of artificial and natural stimuli,Neuron,Volume 109, Issue 10,

2021,Pages 1692-1706.e8, ISSN 0896-6273, https://doi.org/10.1016/j.neuron.2021.03.015.

[7] Tim Gollisch,Features and functions of nonlinear spatial integration by retinal ganglion cells,Journal of Physiology-Paris,Volume 107, Issue 5, 2013, Pages 338-348, ISSN 0928-4257, https://doi.org/10.1016/j.jphysparis.2012.12.001.

[8] Shah, NP, Madugula, S, Grosberg, L, et. al. Optimization of Electrical Stimulation for High-Fidelity Artificial Retina. 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER). 2019; 9:714-718

[9] Grosberg, LE, Ganesan, K, Goetz, GA, et. al. Activation of ganglion cells and axon bundles using epiretinal electrical stimulation. J Neurophysiol. 2017; 118(3):1457-1471

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5596129/

[10] Brackbill, N., Rhoades, C., Kling, A., Shah, N. P., Sher, A., Litke, A. M., & Chichilnisky, E. J. Reconstruction of natural images from responses of primate retinal ganglion cells. *eLife*, *9*, e58516.(2020). https://doi.org/10.7554/eLife.58516

[11] Wandell, B. Foundations of Vision. Sinauer Associates Inc. 1 May 1995, ISBN-13: 978-0878938537 ISBN-10: 0878938532

[12] Orin Packer, David R. Williams, 2 - Light, the Retinal Image, and Photoreceptors, Editor(s): Steven K. Shevell, The Science of Color (Second Edition), Elsevier Science Ltd, 2003, Pages 41-102, ISBN 9780444512512, https://doi.org/10.1016/B978-044451251-2/50003-9.

[13] J. Schwiegerling, *Field Guide to Visual and Ophthalmic Optics*, SPIE Press, Bellingham, WA (2004).

[14] Sterling, P., & Laughlin, S. *Principles of neural design*. (2015).

[15] Nam Phamdo. Theory of Data Compression.Department of Electrical and Computer Engineering, State University of New York. https://faculty.uml.edu//jweitzen/16.548/classnotes/Theory%20of%20Data%20Compression.htm

[16] Kalloniatis, M & Luu, C. Temporal Resolution. WebVision The Organization of the Retina and Visual System. 5 June 2007. https://webvision.med.utah.edu/book/part-viii-psychophysics-of-vision/temporal-resolution/

[17] Artificial vision: what people with bionic eyes see. Conversation. 16 Aug 2017.

https://theconversation.com/artificial-vision-what-people-with-bionic-eyes-see-79758

[18] Huang, G., Gast, T. J., & Burns, S. A. In vivo adaptive optics imaging of the temporal raphe and its relationship to the optic disc and fovea in the human retina. *Investigative ophthalmology & visual science*, *55*(9), 5952–5961. (2014). https://doi.org/10.1167/iovs.14-14893

[19] Rempel, David & Willms, Kirsten & Anshel, Jeffrey & Jaschinski, Wolfgang & Sheedy, James. The Effects of Visual Display Distance on Eye Accommodation, Head Posture, and Vision and Neck Symptoms. Human factors. 49. 830-8. (2007). 10.1518/001872007X230208.

[20] J. E. Farrell, F. Xiao, P. Catrysse, B. Wandell . A simulation tool for evaluating digital camera image quality Proc. SPIE vol. 5294, p. 124-131, Image Quality and System Performance, Miyake and Rasmussen (Eds). (2004).

Appendix

Code Repository:

https://github.com/amrith1/isetbio/tree/dev/arProject