A Single Model Explains both Visual and Auditory Precortical Coding

Honghao Shan,[1] Matthew H. Tong,[1] Garrison W. Cottrell[1*]

[1]Department of Computer Science and Engineering, University of California, San Diego
La Jolla, CA 92093–0404

[*]To whom correspondence should be addressed; E-mail: gary@ucsd.edu.

[1]HS and GWC designed the experiments, HS implemented the models,
HS, MHT, and GWC wrote the manuscript.

March 7, 2024

Precortical neural systems encode information collected by the senses, but the driving principles of the encoding used have remained a subject of debate. We present a model of retinal coding that is based on three constraints: information preservation, minimization of the neural wiring, and response equalization. The resulting novel version of sparse principal components analysis successfully captures a number of known characteristics of the retinal coding system, such as center-surround receptive fields, color opponency channels, and spatiotemporal responses that correspond to magnocellular and parvocellular pathways. Furthermore, when trained on auditory data, the same model learns receptive fields well fit by gammatone filters, commonly used to model precortical auditory coding. This suggests that efficient coding may be a unifying principle of precortical encoding across modalities.

Introduction

Sensory information goes through various forms of processing before it reaches the cerebral cortex. Visual information is transformed into neural signals at the retina, where it passes through retinal ganglion cells that are characterized by their center-surround shaped receptive fields (Enroth-Cugell & Robson, 1966); auditory information, on the other hand, is passed to the brain through the auditory nerve fibers whose filtering properties can be well described by gammatone filters (Kiang, Watanabe, Thomas, & Clark, 1965). Since such peripheral processing prepares the data that the subsequent cortical processing relies on, its functional role has attracted a great deal of attention in the past several decades (Srinivasan, Laughlin, & Dubs, 1982; Field, 1987; Atick & Redlich, 1992; Lewicki, 2002; Vincent, Baddeley, Troscianko, & Gilchrist, 2005; Graham, Chandler, & Field, 2006; Doi & Lewicki, 2007).

Despite intensive research, there are still mysteries concerning the functional role of pre-cortical processing. For example, do different sensory modalities (visual, auditory, somatosensory, etc.) adopt the same computational principles in their pre-cortical stages? Although it is tantalizing to assume so, recent studies suggest otherwise. For example, (Lewicki, 2002) learned gammatone filters from natural sound using independent component analysis (ICA), which was previously applied to natural image patches to learn edge/bar shaped filters resembling the V1 simple cells' receptive fields (Olshausen & Field, 1996; Lewicki & Sejnowski, 2000). Since gammatones model pre-cortical auditory nerve fibers while V1 is a region of cortex, this gives rise to a puzzle: Why would the brain use the same strategy for preprocessing at a *pre-cortical* stage in the auditory pathway and early *cortical* processing in the visual pathway (Olshausen & O'Connor, 2002)?

Questions remain even for peripheral processing in a single modality. Recently, Graham et al. proposed that decorrelation, response equalization, and sparseness form the minimum constraints that must be considered to account for the known linear properties of retinal coding (Graham et al., 2006). This hypothesis is the combination of several previous theories. The response equalization theory hypothesizes that retinal coding seeks a representation that ensures that each neuron has approximately the same average activity level when the animal is presented with natural scenes (Field, 1987; Brady & Field, 1995; Field & Brady, 1997; Brady & Field, 2000). The output decorrelation theory follows the efficient coding principle (Attneave, 1954; Barlow,

1961; Atick & Redlich, 1992). It hypothesizes that retinal coding represents the most efficient coding of the information in the visual domain by capturing the second-order statistical structure of the visual inputs and making the signals from these neurons less correlated. Both of these theories are derivatives of whitening theory, which hypothesizes that retina coding produces a flattened response spectrum for natural visual inputs from a specific range of spatial frequencies (Srinivasan et al., 1982). This whitening theory links the properties of retinal coding with the statistics of natural scenes and is now part of the prevailing view of retinal processing. A third theory suggests that the system is trying to minimize energy usage or wiring cost (Vincent & Baddeley, 2003). Vincent et al. argued that systems that try to minimize energy usage by minimizing wiring give center surround receptive fields. Given these various objectives, it is still not clear what constraints are actually operating in the specification of the retinal coding system. It would be desirable to build a retinal coding model that integrates the different ideas behind these theories and explains the origins of the observed center-surround receptive fields. Ideally, this model should also able to explain the pre-cortical processing of other modalities.

Our model takes into account the following considerations. The retina compresses the approximately 100 million photoreceptor responses into a million ganglion cell responses. Hence the first consideration is that we would like the ganglion cells to retain the maximum amount of information about the photoreceptor responses. If we make the simplifying assumption that ganglion cells respond linearly, then the optimal linear compression technique in terms of reconstruction error is principal components analysis (PCA). One can map PCA into a neural network as in Figure 1(a) (Cottrell, Munro, & Zipser, 1989; Baldi & Hornik, 1989). The weight vectors of each hidden unit in this network each correspond to one eigenvector of the covariance matrix of the data. In standard PCA, there is an ordering to the hidden units, such that the first hidden unit has very high response variance and the last hidden unit has practically no variance, which means the first hidden unit is doing orders of magnitude more work than the last one. The second consideration, then, is that we would like to spread the work evenly among the hidden units. Hence we impose a threshold on the average squared output of the hidden units. As we will see from the simulations, in order to preserve the maximum information, the units all hit this threshold, which equalizes the work. The third consideration is that PCA is profligate with connections - every ganglion cell would have non-zero connections to every

photoreceptor. Hence we also impose a constraint on the connectivity in the network. In this latter constraint we were inspired by the earlier work of (Vincent et al., 2005). They proposed a model of retinal and early cortical processing based on energy minimization and showed that it could create center-surround shaped receptive fields for grayscale images. However, their system sometimes led to cells with two center-surround fields, and the optimization itself was unstable.

These considerations lead to our objective function. Images can be represented as high-dimensional real-valued data; if L photoreceptors are representing the input image, the observed image can be represented as $x \in R^L$. Given input vectors $\mathbf{x} \in R^L$, we seek to find the output responses $\mathbf{s} \in R^M$ (the signal from the retinal ganglion cells) and basis functions $A \in R^{L \times M}$ (the connections from the photoreceptors to the ganglion cells) such that the following objective function is minimized:

$$E = \left\langle \frac{\|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2^2}{2} \right\rangle + \lambda \|\mathbf{A}\|_1 \tag{1}$$

subject to:

$$\left\langle s_i^2 \right\rangle \leq 1 \quad \forall i \tag{2}$$

where $\langle \cdot \rangle$ denotes taking average over all the input samples. The first term in Equation 1 minimizes the reconstruction error and maximizes the information maintained by the encoding. When the sparsity weight λ is small and $L > M$ (i.e., the encoding compresses the information), the reconstruction error reduces to a term that only involves the correlation matrix: $\mathbf{C} = \langle \mathbf{x}\mathbf{x}^t \rangle$ (see supplementary materials). This concurs with the idea that the system is only sensitive to second order statistics. The second constraint minimizes the connections from the photoreceptors to the ganglion cells, incorporating sparsity and an economy of elementary features. The constraint on the average energy of the ganglion cells equalizes the work across the ganglion cells. The system will in fact push this term to the threshold of 1 in order to maintain the maximum information. Thus this objective function integrates three major theories of retinal coding: efficient coding (Attneave, 1954), response equalization (Field, 1987), and the economy of elementary features (Vincent et al., 2005). While this does not directly embody the decorrelation theory (Atick & Redlich, 1992), we have not found that assumption necessary to obtain our results.

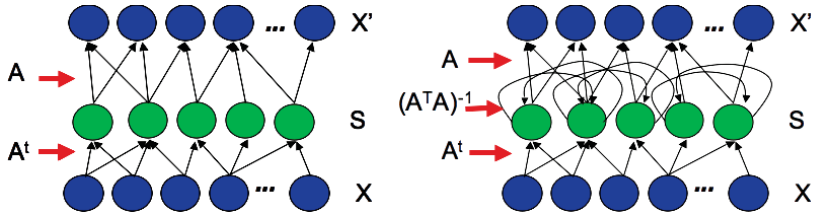
Put another way, the three terms in the objective function determine different aspects of the basis functions (i.e., the columns of $\mathbf{A}$): the reconstruction error determines the subspace that the basis functions span; the output constraint specifies the lengths of the basis functions; and the sparsity penalty rotates the basis functions within the subspace determined by the reconstruction error. In this sense, all three terms, and hence all three theories that they embody, are necessary to fully characterize the retinal coding model. This observation partially agrees with the prediction by (Graham et al., 2006): "we conclude that a minimum of three constraints must be considered to account for the known linear properties – decorrelation, response equalization, and size/sparseness." Our three constraints are efficient coding of information, response equalization, and sparseness of connections.

An important feature of this objective function is that it allows us to derive an efficient algorithm to learn the model parameters, because the revised model turns out to be a particular variation of Sparse PCA (Zou, Hastie, & Tibshirani, 2006) that is reducible to sparse coding (Olshausen & Field, 1996). We can therefore efficiently estimate the parameters of the model and apply it to a larger range of data than has typically been used in the past. A critical insight provided by the mapping to sparse coding is that this model *is* exactly sparse coding, applied to the *transpose* of the data matrix rather than the data matrix itself. This means we can also use our model for dimensionality *expansion* (overcomplete representations) as well as dimensionality reduction, although when doing expansion, we can no longer use the efficient approximation derived in the Supplementary Materials.

In what follows, we show that this simple objective function is able to account for both retinal ganglion cell receptive fields and gammatone filters that have been used to characterize the signals in the auditory nerve.

Results

Going forward, it is important to understand the distinction between *features* and *filters*. The features are the the rows of $\mathbf{A^t}$, and represent the connections between the photoreceptors and the hidden units in Figure 1(b). The filters, on the other hand, correspond to the rows of $\mathbf{W} = (\mathbf{A^t A})^{-\mathbf{1}}\mathbf{A^t}$, the pseudoinverse of $\mathbf{A}$, which correspond to the receptive fields of the ganglion cells that would result from reverse correlation. We visualize the computation as in Figure 1(b): the network receives input from the photoreceptors, and then



(a) PCA Network (b) Sparse PCA Network

Figure 1: Two neural networks that can implement PCA and Sparse PCA. The left hand panel represents a network that performs PCA. The weights (rows of $\mathbf{A^t}$) to each hidden unit (ganglion cell) from the pixels (photoreceptors) represent the coordinates of a unit-length eigenvector of the covariance matrix of the data, so that the activations are the projections onto the eigenvectors. The same weights (transposed) can be used to reconstruct the data (these are not part of the model). The right panel represents Sparse PCA. It is the connections from the pixels that are sparse ($\mathbf{A^t}$). This is followed by recurrent connections that give the center-surround shape. This would be reflected in any recordings of the hidden units, hence the receptive field of a hidden unit is represented by a row of $\mathbf{W} = (\mathbf{A^t A})^{-\mathbf{1}}\mathbf{A^t}$.

there is inhibition of hidden units with similar receptive fields (represented by the recurrent connections $(\mathbf{A^t A})^{-1}$).

Grayscale Images We applied the model to four grayscale image datasets. Results were qualitatively similar across the sets; the results we describe here are from the subset of the Van Hateren natural image set described in (Karklin & Lewicki, 2009). For this simulation, we used 20×20 patches of pixels and reduced them to 100 dimensions. Our SPCA model captures 99.23% of the variance that is captured by standard PCA with 100 eigenvectors retained, while 96.31% of the connection weights (the rows of $\mathbf{A^t}$) are absolute zero; in contrast, none of the connection weights in standard PCA are zero. Figure 8(a) and 8(b) plots the distribution of the connection weights in $\mathbf{A}$ learned by our model versus standard PCA. Each ganglion cell is directly connected to only 3.69% of the input neurons in the 20×20 patch on average. Clearly, this sparsity would be advantageous for a biological system.

The learned elementary features (i.e., the columns of $\mathbf{A}$) are blobs of similar size that tile the 20×20 image patch. The top panel in Figure 2(a) displays 10 features randomly selected from all 100 features. We fit all the features with 2D Gaussians, and plot them as circles in Figure 2(b). The center and the radius of each circle represent the center and twice the standard deviation of the fitted Gaussian. To visualize how well the Gaussians fit the features, we display the first feature in Figure 2(a) and highlight its fitted Gaussian. As shown in the figure, the Gaussians provide a mosaic coverage of the image patch. If we reduce the number of hidden units to 32, the blobs enlarge to cover the image, as shown in Figure 2(d).

The optimal filters (i.e., the rows of $\mathbf{W}$) are center-surround shaped, as shown in the bottom panel of Figure 2(a). The first filter, for example, recovers the weight assigned to the first feature in the top panel. It is tantalizing to think that some of the filters are ON-centered while others are OFF-centered. However, we can switch the signs of the features (and hence the signs of the optimal filters) without changing the model's objective function. Hence our model does not provide insight into the difference between the ON-centered and the OFF-centered cells (Chichilnisky & Kalmar, 2002) beyond the usual explanation that neurons cannot fire both positively and negatively.

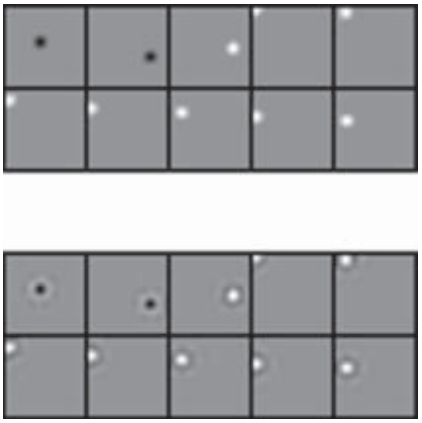
It is interesting to see why a population of Gaussian blob shaped features should give rise to center-surround shaped filters. As shown in Figure 2(c),

each filter is a weighted sum of all the elementary features and can be viewed as the result of a sequence of efforts to recover the contribution of its corresponding feature. Each feature is first applied as a template filter on the image patch to estimate its contribution. However, this estimation is inaccurate because this feature overlaps with its neighboring features. To get a more accurate estimation the contribution from neighboring features must be subtracted. This potentially overcompensates, so get an even more accurate estimation, one must add back the contribution from the features neighboring the features that surround the first feature. This process repeats, moving ever outward. However, the weight reduces quickly for features removed from the first feature, which makes the resulting filter effectively localized and keeps the filter center-surround shaped (for low lambda, some additional ripples can be observed - however, these additional ripples have also been observed in ganglion receptive fields (Dearworth Jr. & Granda, 2002)).

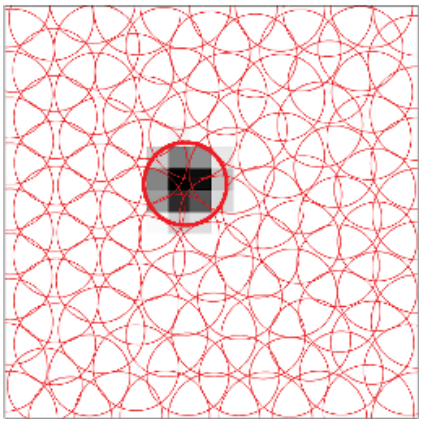
Which aspects of natural scene images give rise to the learned features we observe? To answer this question, we apply our algorithm to white noise images, which contain no statistical structure, and pink noise images, which follow the same $1/f$ power law as natural scene images (Field, 1987) but otherwise contain no structure. Figure 3 displays two example images and some of the learned filters. On white noise images, the learned features are one-pixel image templates; the corresponding filters also only contain one non-zero pixel. That is, the model simply keeps 64 pixels and ignores the other pixels. That's the best it can do with 64 features, because white noise images contain no structure. On pink noise images, we learn essentially the same elementary features (and hence the same filters) as those learned from natural scene images. This result supports the hypothesis that the center-surround shaped filters come from the $1/f$ power spectrum of natural scene images, which agrees with the classic whitening theory (Srinivasan et al., 1982; Field, 1987; Atick & Redlich, 1992).

The sparseness level λ plays an important role in shaping the learned features and filters. As λ increases, the model puts more emphasis on sparse connections at the cost of keeping less information about the inputs. In a biological system, this may occur when the system is on a strict energy budget. Here we check how the learned filters change with larger λ values.

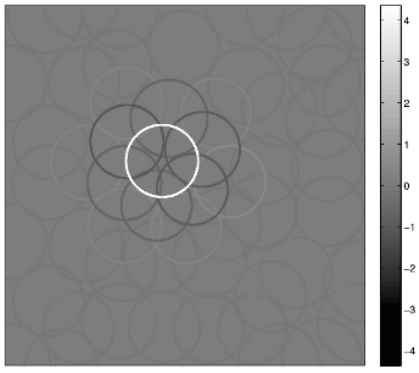
By analyzing the filters in Fourier space, we can plot amplitude at various frequencies, giving a contrast sensitivity function. As shown in Figure 4, with larger λ value, the model becomes less sensitive to low frequency information, but more sensitive to high frequency information. This change matches with



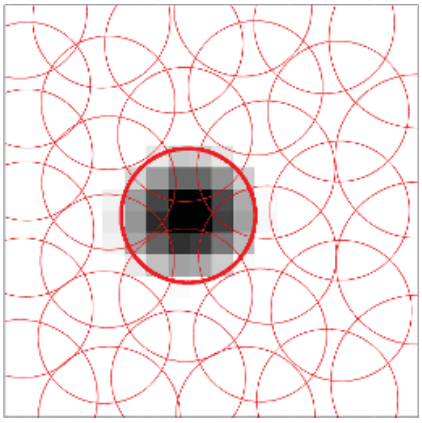
(a) Elementary features and filters



(b) Distribution of 100 elementary features

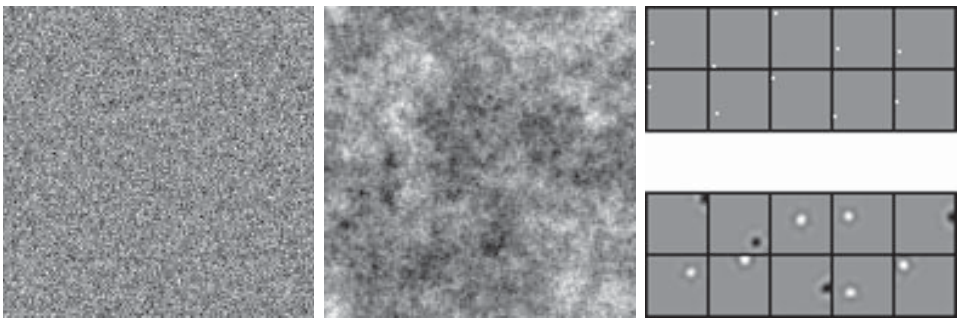


(c) Filter as a weighted sum of features



(d) Distribution of 32 elementary features

Figure 2: Elementary features and their corresponding optimal filters, learned from 20×20 grayscale image patches. In Figure 2(a), the top panel displays 10 features randomly selected from all the 100 features; the bottom panel displays their optimal filters. We fit the features with Gaussian blobs and plot them as circles in Figure 2(b). The radius of each circle represents twice the standard deviation of the fitted Gaussian blob. To help visualize how well the Gaussian blobs fit with the features, we display the first feature in Figure 2(a) and highlight its fitted Gaussian. These Gaussian will become bigger if we use a smaller number of features to "construct" the image patches, as shown in Figure 2(d). Figure 2(c) displays the first filter in Figure 2(a) as a weighted sum of all the 100 features. Each feature is plotted as a circle, as in Figure 2(b) and 2(d). The color of each circle represents the weight assigned to this feature (read the main text for details).



(a) White noise (b) Pink noise (c) Learned filters

Figure 3: Experiments on white noise and pink noise images. Figure 3(a) and 3(b) display images containing white noise and pink noise. The top panel in Figure 3(c) displays 10 filters learned from white noise images; the bottom panel plays the filters learned from pink noise images.

psychophysical studies of contrast sensitivity in children with chronic malnutrition. Compared with normal children, malnourished children are reported to be less sensitive to low spatial frequencies, but slightly more sensitive to high spatial frequencies (dos Santos & Alencar, 2010). This shift of acuity towards high frequencies, as suggested by our result, might due to the effort of the neural system to capture more visual information with a limited neural wiring budget.

Chromatic Images We applied our algorithm to four chromatic image datasets and again found that we learn qualitatively similar features with each. Here, we report the features learned from Kyoto image dataset. Retinal L, M, and S cones are estimated and given as input to the model. The resulting model captures 99.75% of the variance that is captured by an optimal linear model (PCA) with 256 output neurons, with 96.11% of its connections being absolute zero.

Figure 5(a) displays 6 representative features as well as their corresponding filters, learned from chromatic image patches. We visualize the connection strength from three of the filters to the L/M/S channels in Figure 5(b). Among all the 256 learned features, 193 are black/white blobs, 48 are blue/yellow blobs, 15 are red/green blobs. Figure 5(c), 5(d), and 5(e) plot the spatial layout of learned features.

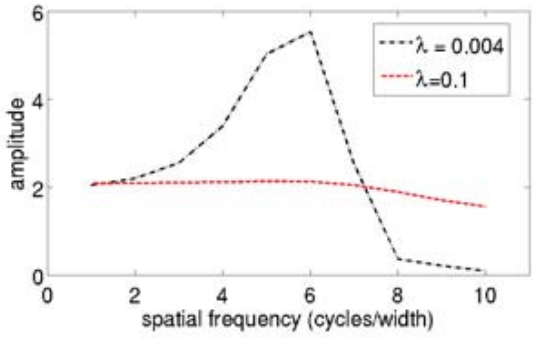
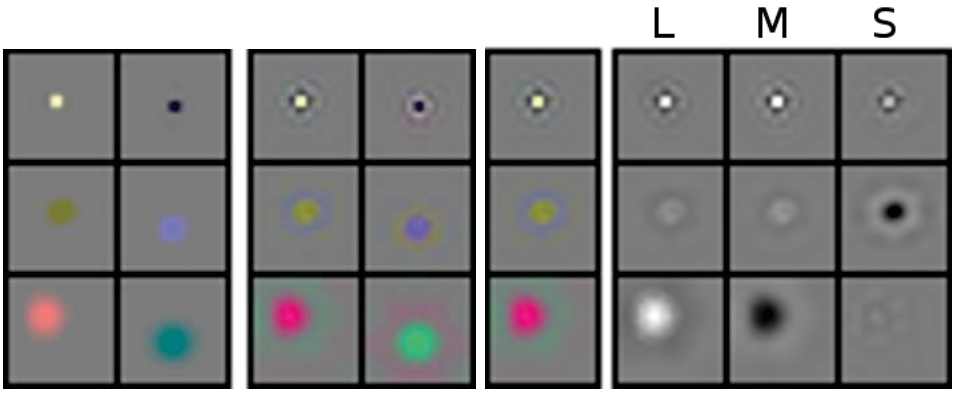


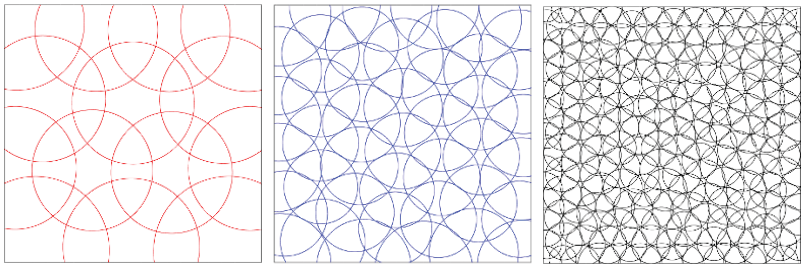
Figure 4: Experiments with increased sparseness level λ. We plot the amplitudes of different frequency component of the filters learned with $\lambda = 0.004$ and $\lambda = 0.1$. As shown in the figure, with an increased λ value, the filter becomes less sensitive to low frequencies, but more sensitive to high frequencies.

The above result replicates the segregation of the spatial channel and the color channel at the retina stage (Calkins & Sterling, 1999). This segregation was explored in previous research that applied information-theoretic methods to natural color spectra/images, such as PCA (Buchsbaum & Gottschalk, 1983; Derrico & Buchsbaum, 1991) and ICA (Tailor, Finkel, & Buchsbaum, 2000; Wachtler, Lee, & Sejnowski, 2001; Lee, Wachtler, & Sejnowski, 2002; Doi, Inui, Lee, Wachtler, & Sejnowski, 2003; Caywood, Willmore, & Tolhurst, 2004). One common observation in these studies is that the learned visual features (eigenvectors or independent components) segregate into black/white, blue/yellow, and red/green opponent structures, either shaped as Fourier basis functions or Gabor kernel functions (Wachtler et al., 2001). These results are similar to what is obtained with ZCA (Zero-component analysis) (Doi et al., 2003), although ZCA has small connections to all of the inputs, since it does not inherently try to minimize connections.

Grayscale Videos To explore the spatio-temporal structure of natural videos, we collected a video dataset of 27 clips from nature documentaries.



(a) Elementary features and filters (b) Connection from the filters to the LMS cones



(c) Red/green (d) Blue/yellow (e) Black/white

Figure 5: Elementary features and the corresponding filters learned from 20×20 chromatic image patches with $\lambda = 0.001$. In Figure 5(a), the left panel displays 6 representative features; the right panel displays their corresponding filters. The features belong to three categories: black/white blobs, blue/yellow blobs, and red/green blobs. The corresponding optimal filters are center-surround shaped, with black/white, blue/yellow, or red/green antagonism. Figure 5(b) plots the connection strength from the filters to the L, M, S cones. Figure 5(c), 5(d), and 5(e) plot the spatial layout of learned features, as we did in Figure 2(b).

Just as a two dimensional image patch can be flattened into a vector of input responses, a three dimensional spatiotemporal patch of video can also be tranformed into a vector. These vectors can then be given to the model as input.

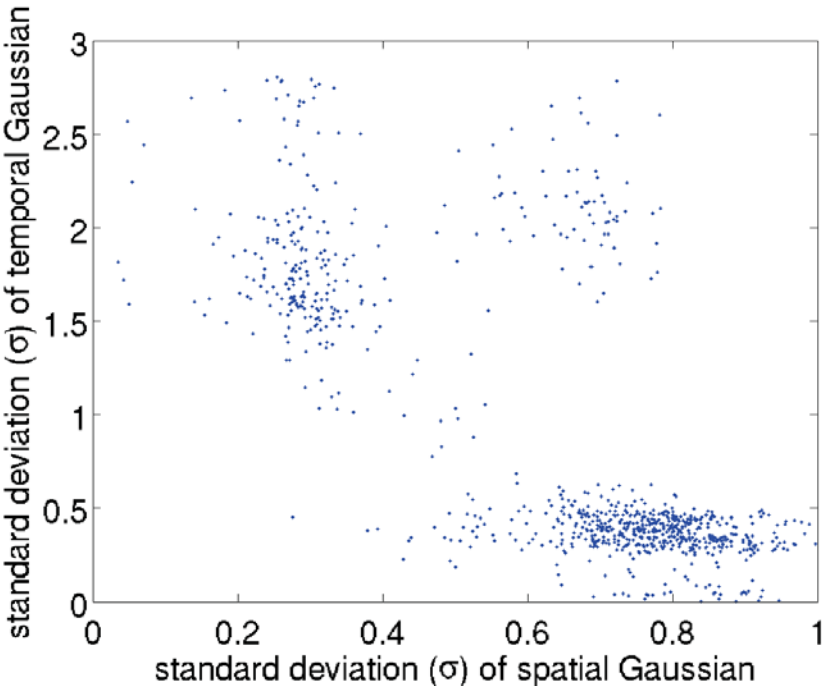
The learned features are black/white blobs whose contrast changes over time. As shown in Figure 6(a), the features can be well fitted to spatio-temporal Gaussians. The corresponding filters are spatially center-surround shaped. Their temporal profile seems to provide an "edge" detector along the temporal axis, which is similar to the temporal profile describing retinal ganglion cells (Pillow et al., 2008). To see the animation file of the learned filters, see `http://cseweb.ucsd.edu/~gary/video_W.gif`.

Another interesting observation is that most of the features segregate into two groups: those with low-spatial and high-temporal frequencies (centered around (0.3,1.75) in the figure), and those with high-spatial and low-temporal frequencies (centered around (.75,0.4) in the figure), as plotted in Figure 6(b). This suggests that the division of ganglion cells into the magno-pathway and the parvo-pathway represents an efficient encoding of the visual environment. This segregation appears to reveal statistical properties of natural videos, instead of coming from our specific algorithm. In fact, we found that this segregation exists even for features learned with standard PCA. As in PCA of static images, however, the features are not biologically plausible.

Sound We applied our algorithm to three sound datasets and get qualitatively similar results on these three datasets. As shown in Figure 7, the learned filters can be well fitted to gammatone filters. Gammatone filters resemble the filtering properties of auditory nerve fibers estimated using the reverse correlation technique from animal such as cats (Carney, 1990) and chinchillas (Recio-Spinoso, Temchin, van Dijk, Fan, & Ruggero, 2005).

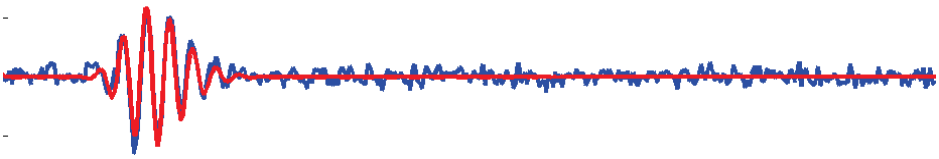
Note that this is (to the best of our knowledge) the first time a non-ICA algorithm has learned gammatone-like filters from sound. Also, since we used the same algorithm for both visual and auditory modalities, this provides an answer to the question posed by Olshausen and O'Connor (2002): "Perhaps an even deeper question is why ICA accounts for neural response properties at the very earliest stage of analysis in the auditory system, whereas in the visual system ICA accounts for the response properties of cortical neurons, which are many synapses removed from photoreceptors." Our model suggests that it is not necessary to use ICA to obtain gammatone filters from sound;

(a) two video filters learned by Sparse PCA (time from left to right)

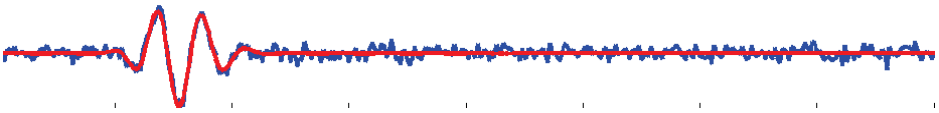


(b) distribution of video features learned by Sparse PCA

Figure 6: Video features segregate into two groups: small, persistent features, and large, brief features.

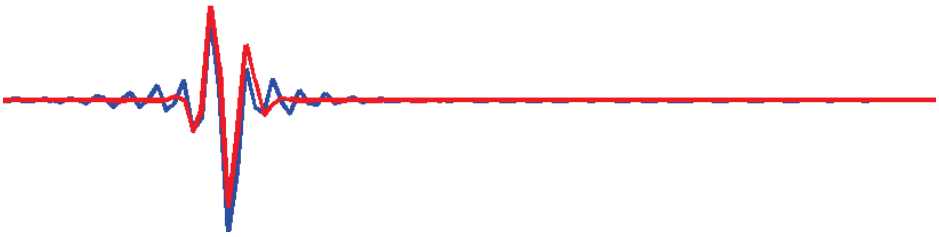


(a) revcor filter from cat



(b) revcor filter from cat

(c) filter learned from the TIMIT dataset



(d) filter learned from the Pittsburgh dataset

Figure 7: Figure 7(a) and 7(b) plot the revcor filters estimated from cat's auditory nerve fibers using the linear reverse correlation technique, as well as the fitted gammatone filter. Figure 7(c) and 7(d) plot filters learned from the TIMIT speech dataset and the Pittsburgh environmental sound dataset respectively. The blue line plots the estimated filter; the red line plots the fitted gammatone filter.

rather, Sparse PCA can account for the receptive fields of neurons at the very earliest stages of analysis in both auditory and visual modalities.

Discussion

We have suggested three principles that can be used to explain precortical encoding: information preservation, minimization of the neural wiring, and response equalization. Each of these principles can be independently justified via evolutionary and energy minimization arguments. Clearly, an organism should try to extract as much relevant information as possible from its environment. As organisms evolve to survive in more enriched environments, which information is relevant becomes more difficult to encode in the genome. As Barlow and Attneave have argued, redundancy reduction for efficient coding is a reasonable response to environmental complexity. Minimizing energy usage suggests constructing the minimal architecture possible, via minimizing wiring, which is a win in terms of both development and daily energy budgets. Finally, equalizing the work (response normalization) results in no single component being crucial to the organism.

The model is closely related to previous theories, but differs in crucial respects. We were inspired by Vincent et al.'s model (Vincent et al., 2005), which also attempted to derive center surround receptive fields by information preservation and minimizing wiring. By including the response normalization constraint, we were able to obtain a more stable algorithm with none of the occasional double receptive fields generated by their model. Unlike many previous models, explicitly decorrelating the outputs of our model was not necessary in order to obtain our results. Hence the model integrates two of the three components suggested as being necessary for any retinal coding model by Graham, Chandler, and Field (Graham et al., 2006) - response equalization and sparseness, but replaces decorrelation with minimization of the neural wiring. The Sparse PCA model has an interesting link to Olshausen's sparse coding model; both models try to minimize the reconstruction error with response equalization, but our model imposes sparseness on the dictionary, while the sparse coding model imposes sparseness on the output. Finally, the model is closely related to that described in (Doi & Lewicki, 2007), where they proposed a model in which the retinal output is a linear transform of the input $\mathbf{s} = \mathbf{W}\mathbf{x}$, ignoring optical blur. The objective of their model is to minimize the difference between the input $\mathbf{x}$ and its

reconstruction **AWx**. Hence their model can be seen as the approximation we take when λ is small and $L > M$ (described in the Supplementary materials). They also include two extra terms during learning, to regularize the average amplitude of the outputs, and to impose sparseness on **W** (instead of **A**). Our version, with sparseness on **A**, leads to a simple interpretation of locally-connected ganglion cells with an inhibitory surround. Also, the convexity properties of our model lead to excellent convergence properties.

We derived an efficient algorithm to learn the model parameters by transforming it into a sparse coding problem. Our approximate algorithm uses only the covariance matrix C, and runs orders of magnitude faster than the exact algorithm, while obtaining results that are less than one percent different in the objective function and learned basis functions. Our approximation works well under the assumption that λ is small and $L > M$. While fast approximation is not a necessary part of a successful model, the speed of computation allows it to be used on more rich data, such as video.

We applied our algorithm to grayscale images, color images, grayscale videos, human speech, and environmental sound, and learned visual and auditory filters that resemble the filtering properties of retinal ganglion cells and auditory nerve fibers. Some of the learned filters are novel. For example, it learns the magno and parvo segregation pathways; and it learns the gammatone filters from natural sound.

Finally, as noted above, our model suggests an answer the question of why ICA gives features corresponding to cortical receptive fields in vision, but seems necessary to obtain gammatone-like filters for sound, which is a pre-cortical level of processing. Our suggestion is that ICA is not necessary to obtain gammatone filters; rather, a PCA algorithm with sparsity and response equalization constraints can result in gammatone filters for sound, while also producing receptive fields similar to peripheral neurons in vision.

Materials and Methods

Model As previously discussed, our model seeks to find the output vectors $\mathbf{s} \in R^M$ (the signal from the retinal ganglion cells) and basis functions $A \in R^{L \times M}$ such that the following objective function is minimized:

$$E = \left\langle \frac{\|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2^2}{2} \right\rangle + \lambda \|\mathbf{A}\|_1 \tag{3}$$

(where $\langle\cdot\rangle$ denotes taking average over all the input samples) subject to the constraint that the average output of each cell $\langle s_i^2\rangle \leq 1$. Upon convergence, the model will satisfy $\langle s_i^2\rangle = 1$ as otherwise the objective function could easily be further reduced.

Vincent et al. interpret $\mathbf{A}$ as the synaptic strength between the input and output neurons; and hence they interpret the sparsity penalty on $\mathbf{A}$ as the desire to minimize the neural wiring cost. From a generative point of view, the columns of $\mathbf{A}$ are the elementary features that the model uses to "construct" the observed inputs. Thus the overall objective function can be understood as capturing most information of the inputs using an economic dictionary of elementary features. However, when $\mathbf{A}$ is fixed and full rank, the optimal output is in fact given by $\mathbf{s}^* = \mathbf{W}\mathbf{x}$, where $\mathbf{W}$ is the pseudoinverse of $\mathbf{A}$ (if we ignore our constraint on the average activation of a cell - an assumption we experimentally verified as yielding a suitable approximate solution). If we apply the linear reverse correlation technique to the model neurons (Chichilnisky, 2001), which recovers the filters transforming the inputs to the outputs, we will get the rows of $\mathbf{W}$ as the model neurons' receptive fields.

The columns of $\mathbf{A}$ and the rows of $\mathbf{W}$ describe different aspects of the model. The columns of $\mathbf{A}$ describe the elementary *features* that the model uses to construct the observed inputs; this forms a kind of visual dictionary. The rows of $\mathbf{W}$, on the other hand, represent the best linear *filters* to recover the weights assigned to the elementary features when generating the observed inputs. Since $\mathbf{W}$ better reflects the properties of cells' receptive fields, we focus primarily on $\mathbf{W}$ throughout the paper.

Our objective function allows us to derive an efficient algorithm to learn the model parameters. Our objective function can be written as:

$$E = \frac{\|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2}{2n} + \lambda\|\mathbf{A}\|_1 \tag{4}$$

where the columns of $\mathbf{X}$ and $\mathbf{S}$ store the input and output vectors; $\|\|_F$ denotes taking the Frobenius norm of a matrix (i.e., the square root of the sum of squared entries); n denotes the number of samples. Our constraint now is that each row of $\mathbf{S}$ is constrained to have L2 norm less than or equal to $\sqrt{n}$.

As shown by (Mairal, Bach, Ponce, & Sapiro, 2010), the Sparse Coding problem (Olshausen & Field, 1996) can be expressed in the matrix factoriza-

tion form as:

$$E = \frac{\|\mathbf{X} - \mathbf{AS}\|_F^2}{2} + \lambda\|\mathbf{S}\|_1 \tag{5}$$

Each column of $\mathbf{A}$ is constrained to have L2 norm less than or equal to q. Sparse PCA can then utilize Sparse Coding algorithms because the objective function in Eq (4) can be re-written as:

$$E = \frac{\|\mathbf{X}^t - \mathbf{S}^t\mathbf{A}^t\|_F^2}{2n} + \lambda\|\mathbf{A}^t\|_1 \tag{6}$$

Now we see that Eq (6) can be symbolically mapped to Eq (5), if we replace $\mathbf{X}^t$ with $\mathbf{X}$, $\mathbf{S}^t$ with $\mathbf{A}$, $\mathbf{A}^t$ with $\mathbf{S}$, with extra care taken to deal with n and $\sqrt{n}$.

The above discussion suggests another interpretation of the retinal coding model: it can be interpreted as removing redundancy between input samples, because Sparse Coding is usually interpreted as removing redundancy between input dimensions (Lewicki & Olshausen, 1999; Lewicki & Sejnowski, 2000). In this sense, Architecture-1 ICA (Bartlett, Lades, & Sejnowski, 1998), which applies ICA to $\mathbf{X}^t$ instead of $\mathbf{X}$, was perhaps the first Sparse PCA algorithm ever proposed.

We then only need to pick some efficient Sparse Coding algorithms to optimize our model parameters. Typically, optimizing the objective function in Eq (5) is factored into two sub-problems: optimizing $\mathbf{A}$ while fixing $\mathbf{S}$, and optimizing $\mathbf{S}$ while fixing $\mathbf{A}$. Both sub-problems are convex optimization problems (this is one reason we impose $\langle s_i^2 \rangle \leq 1$ instead of $\langle s_i^2 \rangle = 1$; otherwise optimizing $\mathbf{S}$ is no longer a convex optimization problem). Recently, it was shown that the coordinate descent algorithm is considerably faster than competing methods for both sub-problems (Friedman, Hastie, & Tibshirani, 2010; Mairal et al., 2010). These algorithms are described in Appendix A.

The above algorithm has a computational complexity that depends on the number of samples n. Here we give an approximate algorithm whose complexity only relies on L and M, the input and output dimensionalities. In our experiments on grayscale images, this reduces the computation time from 37 minutes to 10 seconds, with the learned parameters very close to those learned without approximation. The derivation of the algorithm also helps us to see the connection between this retinal coding model and the output decorrelation theory (Atick & Redlich, 1992).

Our algorithm utilizes two approximations, both of which rely on the condition that λ is small and $L > M$. Under such a condition, the first approximation we use is to calculate the optimal output $\mathbf{s}$ using (See Appendix B for detailed derivations):

$$\mathbf{s}^* \approx \mathbf{W}\mathbf{x} \quad (\mathbf{W} = (\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t) \tag{7}$$

Replacing the above approximation into the original objective function, we get

$$E \approx \frac{\mathrm{Tr}\left((\mathbf{I} - \mathbf{A}\mathbf{W})\mathbf{C}(\mathbf{I} - \mathbf{A}\mathbf{W})^t\right)}{2} + \lambda\|\mathbf{A}\|_1 \tag{8}$$

where $\mathbf{C} = \langle \mathbf{x}\mathbf{x}^t \rangle$, $\mathbf{I}$ denotes the identity matrix, and Tr denotes taking the trace of a matrix. The constraint $\langle s_i^2 \rangle \leq 1$ can be expressed as $\mathrm{diag}\left(\mathbf{W}\mathbf{C}\mathbf{W}^t\right) \leq \mathbf{1}$.

Since $\mathbf{C} = \langle \mathbf{x}\mathbf{x}^t \rangle$ is positive semidefinite, we can factor it using the eigenvalue decomposition $\mathbf{C} = \mathbf{U}\mathbf{V}\mathbf{U}^t$, where $\mathbf{U}$ is a unitary matrix (i.e., $\mathbf{U}\mathbf{U}^t = \mathbf{I}$) containing the eigenvectors as its columns; $\mathbf{V}$ is a diagonal matrix with the eigenvalues on its diagonal. Let $\mathbf{B} = \mathbf{U}\mathbf{V}^{1/2}$, we get $\mathbf{C} = \mathbf{B}\mathbf{B}^t$. Replacing $\mathbf{C} = \mathbf{B}\mathbf{B}^t$ into Eq (8), we get

$$E \approx \frac{\|\mathbf{B}^t - \mathbf{Z}^t\mathbf{A}^t\|_F^2}{2} + \lambda\|\mathbf{A}^t\|_1 \tag{9}$$

where $\mathbf{Z} = \mathbf{W}\mathbf{B}$. The second approximation we use is to relax $\mathbf{Z}$ to a free variable instead of constraining it to $\mathbf{Z} = \mathbf{W}\mathbf{B}$. The constraint becomes that each column of $\mathbf{Z}^t$ should have L2 norm less than or equal to 1.

Now we see that Eq (9) can also be symbolically mapped to Eq (5), if we replace $\mathbf{B}^t$ with $\mathbf{X}$, $\mathbf{Z}^t$ with $\mathbf{A}$, and $\mathbf{A}^t$ with $\mathbf{S}$. Hence the objective function in Eq (9) can also be optimized by efficient Sparse Coding algorithms, such as the coordinate descent algorithms in Appendix A. Since Sparse Coding reduces the redundancy between inputs, the above derivation also brings up another interpretation of the retinal coding model: it can be seen as removing the redundancy between the eigenvectors of $\mathbf{C}$ when λ is small and $L > M$.

Our algorithm efficiently minimizes the objective function. We test two optimization methods: directly optimize the objective function in Eq (6), or first initialize $\mathbf{A}$ by optimizing Eq (9) and then optimize Eq (6). We implement the experiments in single precision on a computer server with Intel Core i7 processors. When we reduce the dimensionality from 400 to

100 and set $\lambda = 0.004$, it takes 37 minutes to directly optimize Eq (6). On the other hand, it takes less than 10 seconds to initialize $\mathbf{A}$ by optimizing Eq (9) and another 40 seconds to optimize Eq (6). As shown in Figure 10(a), our approximate method efficiently minimizes the objective function. In fact, after $\mathbf{A}$ is initialized with the approximate method, further optimizing the objective function with the direct method only causes a less than 0.1% change of the objective function, on average the weights in $\mathbf{A}$ are only changed by 0.4%, and no visible difference in the features and filters can be observed. The closeness of the approximation means that in later experiments (color images, video, and sound) the approximation provided by Eq (9) is used without additional optimization.

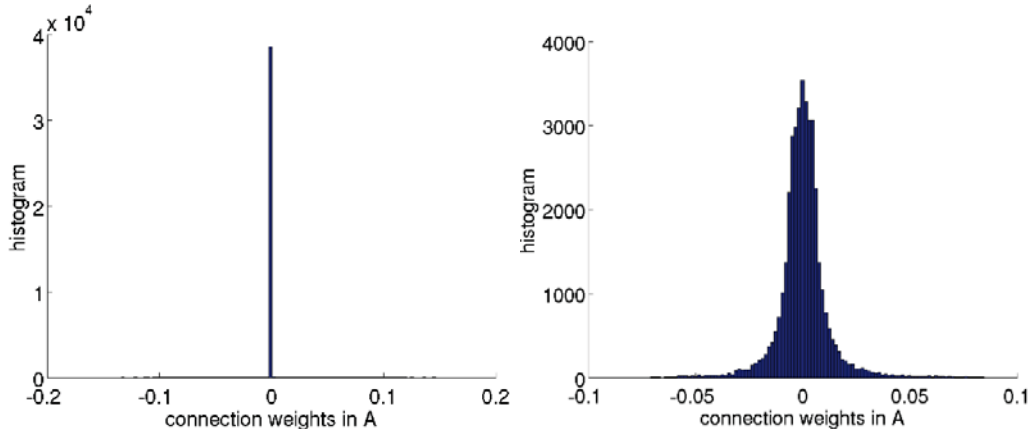
Grayscale Images The four image sets used were van Hateren natural images (van Hateren & van der Schaaf, 1998), Kyoto natural images (Doi et al., 2003), Berkeley segmentation dataset (Martin, Fowlkes, Tal, & Malik, 2001), and Caltech-256 object category dataset (Griffin, Holub, & Perona, 2007). We didn't observe any qualitative difference between the features learned from these datasets. The features reported are obtained using van Hateren natural images. We use a subset of it selected by (Karklin & Lewicki, 2009), which contains 110 1536×1024 grayscale images.

For each image, we discard two pixels off the image borders, normalize the pixel values to $[0, 1]$, then apply a nonlinear function that simulates the cone processing (Karklin & Lewicki, 2009):

$$x = 1 - \exp(-k \cdot x) \tag{10}$$

where x denotes the pixel value, and k is selected for each image such that its average pixel value equals 0.5 after the nonlinearity. This nonlinearity does not seem to alter the features learned by the retinal coding model, but might help to expose higher-order statistical structure (Karklin & Lewicki, 2009). Then we randomly sample 1000 20×20 image patches from each image. Figure 10 compares the distributions of connection weights between PCA and SPCA.

Chromatic Images We applied our algorithm to four chromatic image datasets: the Kyoto natural image dataset (Doi et al., 2003), the McGill color image dataset (Olmos & Kingdom, 2004), the Berkeley segmentation dataset (Martin et al., 2001), and the Caltech-256 object category



(a) Distribution of connection weights using SPCA

(b) Distribution of connection weights using PCA

Figure 8: Comparison of SPCA and PCA connection weights on grayscale image patches. Figure 8(a) and 8(b) plot the distribution of the connection weights in $\mathbf{A}$ learned by our model versus those learned by standard PCA.

dataset (Griffin et al., 2007). From these datasets we learn qualitatively similar features. Below we report the features learned from Kyoto image dataset, which contains 62 1000×1280 pixel chromatic images of natural scenes.

First, we estimate the retinal L, M, S cones' responses to those images. The original images are stored in sRGB color representation (Stokes, Anderson, Chandrasekar, & Motta, 1996). We normalize the pixel values of each image to $[0, 1]$, then transform the image to the CIE XYZ color space (T. Smith & Guild, 1931), from which we estimate the LMS cone responses following the CIECAM02 color appearance model (Fairchild, 2001). In this manner, we can roughly estimate how the retinal L, M, and S cones would respond when presented with the image content.

After that, we apply the cone nonlinearity in Eq (10) to the estimated LMS cone responses. Then we extract all 20 by 20 image patches, and estimate the matrix $\mathbf{C}$. We apply the retinal coding model using the approximation in Eq (9) to reduce the dimensionality from 1200 to 256 with $\lambda = 0.002$. The resulting model captures 99.75% of the variance captured by an optimal linear model (i.e., standard PCA) with 256 output neurons, with 96.11% of its connections being absolute zero.

Grayscale Videos To explore the spatio-temporal structure of natural videos, we collected a video dataset of 27 clips from YouTube (see online supporting material for their URLs). All the videos are from natural history shows from the BBC World Wide channel (`www.youtube.com/user/BBCWorldwide`). In order to have a realistic sample of natural videos, we eliminate those that are obviously unnatural, such as walking dinosaurs, cavemen, or pigeon-mounted cameras. We calculated the power spectrum of the videos in order to eliminate interlaced-format videos, which have an ellipse-shaped power spectrum elongated in the horizontal direction. We initially applied our algorithm to van Hateren video dataset (van Hateren & Ruderman, 1998). The learned features and filters are qualitatively similar to those reported below, except that some of the learned features are ellipse or even bar shaped and are elongated along the horizontal direction. These features most likely result from the fact that the original videos are interlaced (i.e., recording only the odd numbered lines in one frame, and the even numbered lines in the next frame). Although the videos were de-interlaced by block averaging with $2 \times 2 \times 2$ (van Hateren, personal communication), their spatial power spectrum is still ellipse-shaped with more energy along the horizontal direction.

The original YouTube videos are in color. We transform them to grayscale videos using the method described by the Matlab function rgb2gray, normalize the pixel values between $[0, 1]$, and apply the cone nonlinearity in Eq 10. Then we estimate the correlation matrix $\mathbf{C} = \mathbf{x}\mathbf{x}^t$ for all the $12 \times 12 \times 12$ video cubes (with the local mean removed - keeping in the local mean did not substantially change the qualitative results, but it did seem to produce slightly noisier filters), and apply our algorithm to $\mathbf{C}$ using the approximate method.

Sound We applied our algorithm to three sound datasets: the Pittsburgh natural sounds dataset (E. C. Smith & Lewicki, 2006), the TIMIT speech dataset (Lamel, Kassel, & Seneff, 1986), and rainforest mammal vocalization dataset (Emmons, Whitney, & Ross, n.d.). The Pittsburgh natural sounds dataset contains 48 recordings of natural sound recorded around the Pittsburgh region, including ambient sounds (such as rain, wind, and streams) and quick acoustic events (such as snapping twigs, breaking wood, and rock impacts). The TIMIT speech dataset contains English speech from 630 speakers, with each person speaking 10 sentences. The rainforest mammal vocal-

ization dataset contains the characteristic sounds of 109 species of rainforest mammals, such as primates, anteaters, bats, jaguars, and manatees.

For each dataset, each recording is re-sampled at 16 kHz. We normalize the maximum amplitude of each recording to 1, take all the segments formed using a sliding window of 128 sample points (about a 8 millisecond window), then estimate the matrix $\mathbf{C}$ for the sound segments. We then apply the approximate method to learn the features and filters.

References

Atick, J. J., & Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation*, *4*(2), 196–210.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3), 183–193.

Baldi, P., & Hornik, K. (1989). Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks*, *2*, 53–58.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 217–234). Cambridge, MA, USA: MIT Press.

Bartlett, M. S., Lades, M., & Sejnowski, T. J. (1998, January). Independent component representations for face recognition. *Proceedings of the SPIE Symposium on Electonic Imaging: Science and Technology; Conference on Human Vision and Electronic Imaging III*, *3299*.

Brady, N., & Field, D. J. (1995). What's constant in contrast constancy? the effects of scaling on the perceived contrast of bandpass patterns. *Vision Research*, *35*(6), 739–756.

Brady, N., & Field, D. J. (2000). local contrast in natural images: normalisation and coding efficiency. *Perception*, *29*(9), 1041–1055.

Buchsbaum, G., & Gottschalk, A. (1983). Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proceedings of the Royal Society B: Biological Sciences*, *220*(1218), 89–113.

Calkins, D. J., & Sterling, P. (1999, October). Evidence that circuits for spatial and color vision segregate at the first retinal synapse. *Neuron*, *24*, 313–321.

Carney, L. H. (1990). Sensitivities of cells in anteroventral cochlear nucleus of cat to spatiotemporal discharge patterns across primary afferents. *Journal of neurophysiology*, *64*(2), 437–456.

Caywood, M. S., Willmore, B., & Tolhurst, D. J. (2004, June). Independent components of color natural scenes resemble V1 neurons in their spatial and color tuning. *Journal of Neurophysiology*, *91*(6), 2859–2873.

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network: Computation in Neural Systems*, *12*(2), 199–213.

Chichilnisky, E. J., & Kalmar, R. S. (2002). Functional asymmetries in ON and OFF ganglion cells of primate retina. *Journal of Neuroscience*, *22*(7), 2737–2747.

Cottrell, G. W., Munro, P., & Zipser, D. (1989). Image compression by back propagation: An example of extensional programming. In N. E. Sharkey (Ed.), *Models of cognition: A review of cognitive science* (pp. 208–240). Norwood, New Jersey: Ablex.

Dearworth Jr., J. R., & Granda, A. M. (2002). Multiplied functions unify shapes of ganglion-cell receptive fields in retina of turtle. *Journal of Vision*, *2*(3).

Derrico, J. B., & Buchsbaum, G. (1991). A computational model of spatiochromatic image coding in early vision. *Journal of Visual Communication and Image Representation*, *2*(1), 31–38.

Doi, E., Inui, T., Lee, T.-W., Wachtler, T., & Sejnowski, T. J. (2003). Spatiochromatic receptive field properties derived from information-theoretic analyses of cone mosaic responses to natural scenes. *Neural Computation*, *15*(2), 397–417.

Doi, E., & Lewicki, M. S. (2007). A theory of retinal population coding. In *Advances in neural information processing systems* (Vol. 19, pp. 353–360). Cambridge, MA, USA: MIT Press.

dos Santos, N. A., & Alencar, C. C. G. (2010, August). Early malnutrition diffusely affects children contrast sensitivity to sine-wave gratings of different spatial frequencies. *Nutritional Neuroscience*, *13*(4), 189–194.

Emmons, L. H., Whitney, B. M., & Ross, D. L. (n.d.). *Sounds of neotropical rainforest mammals [audio cd] (library of natural sounds, cornell laboratory of ornithology, ithaca, new york, 1997).*

Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology*, *187*(3), 517–552.

Fairchild, M. D. (2001). A revision of CIECAM97s for practical applications. *Color Research & Application*, *26*(6), 418–427.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of American, A*, *4*(12), 2379–2394.

Field, D. J., & Brady, N. (1997). Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research*, *37*(23), 3367–3383.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*, 1–22.

Graham, D. J., Chandler, D. M., & Field, D. J. (2006). Can the theory of "whitening" explain the center-surround properties of retinal ganglion cell receptive fields? *Vision Research*, *46*(18), 2901–2913.

Griffin, G., Holub, A., & Perona, P. (2007). *Caltech-256 object category dataset* (Tech. Rep. No. 7694). California Institute of Technology. Retrieved from `http://authors.library.caltech.edu/7694`

Jolliffe, I. T. (2002). *principal component analysis*. Spring-Verlag.

Karklin, Y., & Lewicki, M. S. (2009). Emergence of complex cell properties by learning to generalize in natural scenes. *Nature*, *457*, 83–86.

Kiang, N. Y.-S., Watanabe, T., Thomas, E. C., & Clark, L. F. (1965). *Discharge patterns of single fibers in the cat's auditory nerve*. MIT Press Cambridge.

Lamel, L. F., Kassel, R. H., & Seneff, S. (1986). Speech database development: design and analysis of the acoustic-phonetic corpus. In *Proceedings of the DARPA speech recognition workshop* (pp. 100–109).

Lee, T.-W., Wachtler, T., & Sejnowski, T. J. (2002). Color opponency constitutes a sparse representation for the chromatic structure of natural scenes. *Vision Research*, *42*(17), 2095–2103.

Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature Neuroscience*, *5*(4), 356–363.

Lewicki, M. S., & Olshausen, B. A. (1999). A probabilistic framework for the adaptation and comparison of image codes. *Journal of the Optical Society of America A*, *16*(7), 1587–1601.

Lewicki, M. S., & Sejnowski, T. J. (2000). Learning overcomplete representations. *Neural Computation*, *12*(2), 337–365.

Mairal, J., Bach, F., Ponce, J., & Sapiro, G. (2010). Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, *11*, 10–60.

Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001, July). A database

of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the 8th international conference on computer vision* (Vol. 2, pp. 416–423).

Olmos, A. A., & Kingdom, F. A. A. (2004). *Mcgill calibrated colour image database.* `http://tabby.vision.mcgill.ca`.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.

Olshausen, B. A., & O'Connor, K. N. (2002). A new window on sound. *Nature*, *5*(4), 292–294.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., & Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, *454*(7207), 995–999.

Recio-Spinoso, A., Temchin, A. N., van Dijk, P., Fan, Y.-H., & Ruggero, M. A. (2005). Wiener-kernel analysis of responses to noise of chinchilla auditory-nerve fibers. *Journal of neurophysiology*, *93*(6), 3615–3634.

Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, *439*(7079), 800–805.

Smith, T., & Guild, J. (1931). The CIE colorimetric standards and their use. *Transactions of the Optical Society*, *33*, 73–134.

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *216*(1205), 427–459.

Stokes, M., Anderson, M., Chandrasekar, S., & Motta, R. (1996). A standard default color space for the Internet–sRGB. *Microsoft and Hewlett-Packard Joint Report*.

Tailor, D. R., Finkel, L. H., & Buchsbaum, G. (2000). Color-opponent receptive fields derived from independent component analysis of natural images. *Vision Research*, *40*(19), 2671–2676.

van Hateren, J. H., & Ruderman, D. L. (1998). Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proceedings of the Royal Society B: Biological Sciences*, *265*(1412), 2315–2315.

van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society B: Biological Sciences*,

265(1394), 359–366.

Vincent, B. T., & Baddeley, R. J. (2003). Synaptic energy efficiency in retinal processing. *Vision Research*, *43*(11), 1283–1290.

Vincent, B. T., Baddeley, R. J., Troscianko, T., & Gilchrist, I. D. (2005). Is the early visual system optimised to be energy efficient? *Network: Computation in Neural Systems, special issue on Sensory Coding and the Natural Environment*, *16*(2/3), 1283–1290.

Wachtler, T., Lee, T.-W., & Sejnowski, T. J. (2001). Chromatic structure of natural scenes. *Journal of the Optical Society of America A*, *18*(1), 65–77.

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of computational and graphical statistics*, *15*(2), 265–286.

Acknowledgements

We thank Lingyun Zhang, Wensong Xu, and Chris Kanan for helpful discussions, Ben Vincent for sharing the source code of his model, Yan Karklin for sharing his image preprocessing code, Eizaburo Doi for sharing his code to calculate LMS cone representations, Vivienne Ming and Mike Lewicki for providing the auditory data, Malcolm Slaney for suggesting the experiments on noise images, and Terry Sejnowski for sharing computational resources.

Supporting Online Material for

A single model explains both visual and auditory precortical coding

Honghao Shan, Matthew H. Tong, Garrison W. Cottrell

A Coordinate Descent Algorithms

Suppose we want to optimize an objective function $f(\mathbf{x})$, where $\mathbf{x}$ is a high dimensional vector. To do so, the coordinate descent algorithm cyclically optimizes each dimension of $\mathbf{x}$. Each time, it optimizes one dimension of $\mathbf{x}$ while fixing the other dimensions. Once this dimension is optimized, the algorithm optimizes another dimension. This process repeats until the objective function can no longer be optimized.

Here we list the coordinate descent algorithm for optimizing the Sparse Coding objective function:

$$E = \left\langle \frac{1}{2}\|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2^2 + \lambda\|\mathbf{s}\|_1 \right\rangle \tag{11}$$

Each column of $\mathbf{A}$ is constrained to have L2 norm less than or equal to q.

Optimizing s While Fixing A When we try to optimize the i-th coordinate of $\mathbf{s}$, with $\mathbf{A}$ and all the other coordinates of $\mathbf{s}$ being fixed, the optimal s_i is given by (Friedman et al., 2010):

$$y = \mathbf{a}_i^t\mathbf{x} - \sum_{j\neq i}\mathbf{a}_i^t\mathbf{a}_j s_j \tag{12}$$

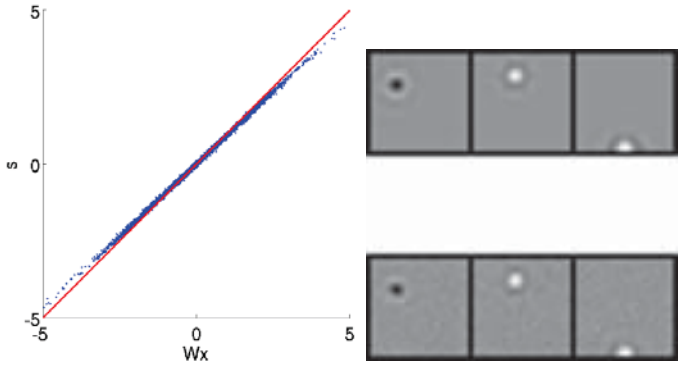
$$s_i^* = \begin{cases} (y-\lambda)/\|\mathbf{a}_i\|_2^2, & \text{if } y > \lambda; \\ (y+\lambda)/\|\mathbf{a}_i\|_2^2, & \text{if } y < -\lambda; \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

where $\mathbf{a}_i$ denotes the i-th column of $\mathbf{A}$.

Optimizing A While Fixing s When we try to optimize the i-th column of $\mathbf{A}$, with $\mathbf{s}$ and all the other columns of $\mathbf{A}$ being fixed, the optimal $\mathbf{a}_i$ is given by (Mairal et al., 2010):

$$\mathbf{u} = \langle s_i\mathbf{x}\rangle - \sum_{j\neq i}\mathbf{a}_j\langle s_j s_i\rangle \tag{14}$$

$$\mathbf{a}_i^* = \frac{\mathbf{u}}{\max(\langle s_i^2\rangle, \|\mathbf{u}\|_2/q)} \tag{15}$$



(a) **Wx** versus **s**

(b) Linear reverse correlation

Figure 9: **W** approximates the model neurons' receptive fields. In Figure 9(a), we compare **Wx** with the optimal **s** inferred using the direct method during the experiments on grayscale image patches. If $\mathbf{s} = \mathbf{Wx}$ holds exactly, all the points should lie on the red solid line. Once we have learned the optimal **A** on grayscale image patches, we feed the model with white noise inputs and infer their optimal outputs, then use $\langle \mathbf{x}\mathbf{s}^t \rangle$ to estimate the model neurons' receptive fields. The top panel in Figure 9(b) displays three rows of **W**; the bottom panel displays the estimated receptive fields for the corresponding neurons.

B Derivation of the approximate algorithm

The revised retinal coding model aims to minimize the following objective function:

$$E = \left\langle \frac{\|\mathbf{x} - \mathbf{As}\|_2^2}{2} \right\rangle + \lambda \|\mathbf{A}\|_1 \tag{16}$$

subject to the constraint that

$$\left\langle s_i^2 \right\rangle \leq 1 \quad \text{for every } i \tag{17}$$

Below we show how to approximate the above objective function when λ is small and $L > M$ (i.e., when we are reducing the data dimensionality).

The First Approximation Let's first check the functional roles of the three terms in the objective function: the reconstruction error, the sparsity

penalty, and the constraint.

If the objective function only contains the reconstruction error term, the model is reduced to the standard PCA problem (Jolliffe, 2002). The optimal basis functions (i.e., the columns of $\mathbf{A}$) should span the subspace spanned by the eigenvectors of $\langle \mathbf{x}\mathbf{x}^t \rangle$ with top eigenvalues. The optimal outputs are given by $\mathbf{s}^* = \mathbf{W}\mathbf{x}$, where $\mathbf{W}$ is the pseudo-inverse of $\mathbf{A}$. Changing the basis functions' individual directions and lengths within the subspace won't change the reconstruction error, because for any full-rank matrix $\mathbf{G} \in R^{M \times M}$ we have

$$\mathbf{A}\mathbf{s} = (\mathbf{A}\mathbf{G})(\mathbf{G}^{-1}\mathbf{s}) \tag{18}$$

That is, for any $\mathbf{A}_{new} = \mathbf{A}\mathbf{G}$ which spans the same subspace as $\mathbf{A}$ but with different individual lengths and directions, we can find $\mathbf{s}_{new} = \mathbf{G}^{-1}\mathbf{s}$ such that the reconstruction error remains to be the minimum.

The objective function is not changed by adding the constraint $\langle s_i^2 \rangle \leq 1$. For any value of $\mathbf{A}$ and $\mathbf{s}$, we can always divide s_i (the i-th coordinate of $\mathbf{s}$) and multiply $\mathbf{a}_i$ (the i-th column of $\mathbf{A}$) with some value α to satisfy the constraint without changing the reconstruction error. In other words, the constraint term only specifies the lengths of the basis functions.

The sparsity penalty $\|\mathbf{A}\|_1$ will shrink the basis functions' lengths and rotate their directions. However, when $\lambda > 0$ is sufficiently small such that the subspace that the basis functions span is mainly determined by the reconstruction error, the sparsity penalty will only serve to rotate the basis functions within the subspace determined by the reconstruction error.

Our first approximation utilizes the above analysis. When λ is small and $L > M$, we use $\mathbf{s}^* \approx \mathbf{W}\mathbf{x}$, where $\mathbf{W}$ is the pseudo-inverse of $\mathbf{A}$. Substituting it into the objective function in Eq (3), we get:

$$\begin{aligned}
E &\approx \left\langle \frac{\|\mathbf{x} - \mathbf{A}(\mathbf{W}\mathbf{x})\|_2^2}{2} \right\rangle + \lambda\|\mathbf{A}\|_1 && (19)\\
&= \left\langle \frac{\mathrm{Tr}\left((\mathbf{x} - \mathbf{A}\mathbf{W}\mathbf{x})(\mathbf{x} - \mathbf{A}\mathbf{W}\mathbf{x})^t\right)}{2} \right\rangle + \lambda\|\mathbf{A}\|_1 && (20)\\
&= \frac{\mathrm{Tr}\left((\mathbf{I} - \mathbf{A}\mathbf{W})\left\langle \mathbf{x}\mathbf{x}^t \right\rangle (\mathbf{I} - \mathbf{A}\mathbf{W})^t\right)}{2} + \lambda\|\mathbf{A}\|_1 && (21)\\
&= \frac{\mathrm{Tr}\left((\mathbf{I} - \mathbf{A}\mathbf{W})\mathbf{C}(\mathbf{I} - \mathbf{A}\mathbf{W})^t\right)}{2} + \lambda\|\mathbf{A}\|_1 \quad (\text{let } \mathbf{C} = \langle \mathbf{x}\mathbf{x}^t \rangle) && (22)
\end{aligned}$$

where $\mathbf{I}$ denotes the identity matrix, Tr denotes taking the trace of a matrix.

The constraint can now be approximated as:

$$\left\langle s_i^2 \right\rangle \approx \left\langle (\mathbf{w}_i \mathbf{x})(\mathbf{w}_i \mathbf{x})^t \right\rangle = \mathbf{w}_i \left\langle \mathbf{x}\mathbf{x}^t \right\rangle \mathbf{w}_i^t \leq 1, \quad \text{or} \quad \text{diag}(\mathbf{W}\mathbf{C}\mathbf{W}^t) \leq \mathbf{1} \quad (23)$$

where $\mathbf{w}_i$ denotes the i-th row of $\mathbf{W}$, and $\mathbf{1}$ denotes a vector of 1's. Hence, when λ is small and $L > M$, the model mainly serves to capture the second-order statistical structure of the inputs.

The Second Approximation Since $\mathbf{C} = \langle \mathbf{x}\mathbf{x}^t \rangle$ is positive semidefinite, we can factor it using the eigenvalue decomposition $\mathbf{C} = \mathbf{U}\mathbf{V}\mathbf{U}^t$, where $\mathbf{U}$ is a unitary matrix (i.e., $\mathbf{U}\mathbf{U}^t = \mathbf{I}$) containing the eigenvectors as its columns; $\mathbf{V}$ is a diagonal matrix with the eigenvalues on its diagonal. Let $\mathbf{B} = \mathbf{U}\mathbf{V}^{1/2}$, we get $\mathbf{C} = \mathbf{B}\mathbf{B}^t$. Substituting this into the objective function in Eq (22), yields

$$\begin{aligned} E &= \frac{\text{Tr}\left((\mathbf{I} - \mathbf{A}\mathbf{W})\mathbf{B}\mathbf{B}^t(\mathbf{I} - \mathbf{A}\mathbf{W})^t\right)}{2} + \lambda\|\mathbf{A}\|_1 && (24) \\ &= \frac{\text{Tr}\left((\mathbf{B} - \mathbf{A}\mathbf{W}\mathbf{B})(\mathbf{B} - \mathbf{A}\mathbf{W}\mathbf{B})^t\right)}{2} + \lambda\|\mathbf{A}\|_1 && (25) \\ &= \frac{\|\mathbf{B} - \mathbf{A}\mathbf{W}\mathbf{B}\|_F^2}{2} + \lambda\|\mathbf{A}\|_1 && (26) \\ &= \frac{\|\mathbf{B} - \mathbf{A}\mathbf{Z}\|_F^2}{2} + \lambda\|\mathbf{A}\|_1 \quad (\text{let } \mathbf{Z} = \mathbf{W}\mathbf{B}) && (27) \end{aligned}$$

The constraint becomes that each row of $\mathbf{Z}$ should have L2 norm less than or equal to 1:

$$\text{diag}(\mathbf{W}\mathbf{C}\mathbf{W}^t) = \text{diag}(\mathbf{W}\mathbf{B}\mathbf{B}^t\mathbf{W}^t) = \text{diag}(\mathbf{Z}\mathbf{Z}^t) \leq \mathbf{1} \quad (28)$$

Our second approximation is to relax $\mathbf{Z}$ to a free variable instead of constraining it to $\mathbf{Z} = \mathbf{W}\mathbf{B}$ because when λ is small and $L > M$, this free variable will converge to $\mathbf{Z} \approx \mathbf{W}\mathbf{B}$ following the same analysis in our first approximation. Now the objective function can be written as:

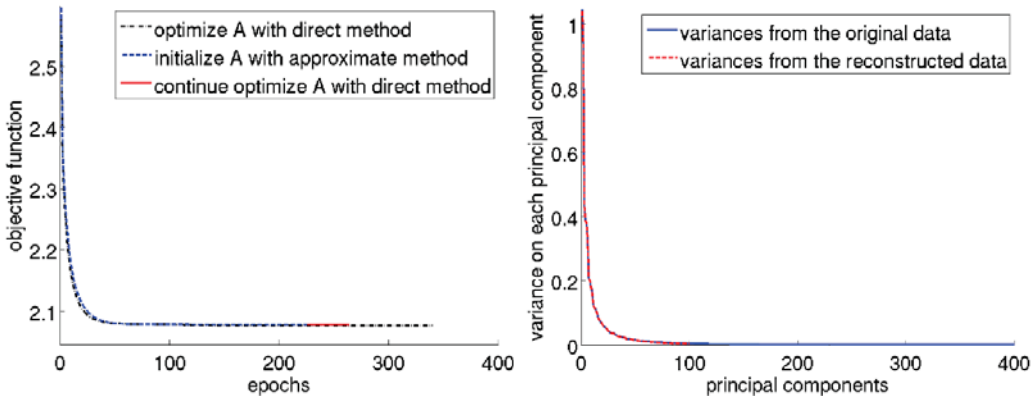
$$E = \frac{\|\mathbf{B}^t - \mathbf{Z}^t\mathbf{A}^t\|_F^2}{2} + \lambda\|\mathbf{A}^t\|_1 \quad (29)$$

Each column of $\mathbf{Z}^t$ should have L2 norm less than or equal to 1. We see that this objective function can also be symbolically mapped to Eq (5). Hence its parameters can be optimized by efficient Sparse Coding algorithms.

In the grayscale image experiment, we verified the efficacy of this approximation. We tested two optimization methods: directly optimizing the objective function in Eq (6), or first initialize $\mathbf{A}$ by optimizing Eq (9) and then optimize Eq (6). We implemented the experiments in single precision on a computer server with Intel Core i7 processors. When we reduce the dimensionality from 400 to 100 and set $\lambda = 0.004$, it takes 37 minutes to directly optimize Eq (6). On the other hand, it takes less than 10 seconds to initialize $\mathbf{A}$ by optimizing Eq (9) and another 40 seconds to optimize Eq (6). As shown in Figure 10(a), our approximate method efficiently minimizes the objective function. In fact, after $\mathbf{A}$ is initialized with the approximate method, further optimizing the objective function with the direct method only causes a less than 0.1% change of the objective function, and on average, the weights in $\mathbf{A}$ are only changed by 0.4%. No visible difference in the features and filters can be observed. Because of the closeness of the approximation, in later experiments (color images, video, and sound), we used the approximation provided by Eq (9) without additional optimization.

C URLs of Video Clips

www.youtube.com/watch?v=cMIRwCNvI94 www.youtube.com/watch?v=J7eRGHVx3p0 www.youtube.com/watch?v=8R1g0t00vGM www.youtube.com/watch?v=K61FRGpvfwE www.youtube.com/watch?v=M-nXN5SGmhw www.youtube.com/watch?v=tOn2RhH36Mc www.youtube.com/watch?v=gc9jFmkjizQ www.youtube.com/watch?v=ZiW96Uci624 www.youtube.com/watch?v=1YQrLPW5DdY www.youtube.com/watch?v=xKksJ3fvB1Q www.youtube.com/watch?v=43id_NRajDo www.youtube.com/watch?v=NRWehNVSAlA www.youtube.com/watch?v=oJ-KzdRsQC4 www.youtube.com/watch?v=VuMRDZbrdXc www.youtube.com/watch?v=2rlZVtKKWnk www.youtube.com/watch?v=aIQB0NFcFog www.youtube.com/watch?v=B71T_GpA2AM www.youtube.com/watch?v=XB-8Grn6sRo www.youtube.com/watch?v=JxrIWShNPko www.youtube.com/watch?v=gVjqL-9Fh3E www.youtube.com/watch?v=yKKabd3W904 www.youtube.com/watch?v=4ZFoqh8PQ88 www.youtube.com/watch?v=NR3Z4p5hspI www.youtube.com/watch?v=RB9uzMjiYSQ www.youtube.com/watch?v=a7XuXi3mqYM www.youtube.com/watch?v=PBrStxuOJbs www.youtube.com/watch?v=u6ouWOGJk5E



(a) Objective function during learning (b) Preserved variance

Figure 10: Experiments on grayscale image patches. Figure 10(a) plots how the objective function changes during learning. The black dash-dotted line plots the objective function using the direct optimization method; the blue dashed line plots the objective function when we use the approximate method to find $\mathbf{A}$; the red solid line plots the objective function when we further fine tune the parameters using the direct method after initializing A with the approximate method. Figure 10(b) plots the eigenvalues of $\mathbf{C}$ from the original data (the blue solid line) versus those from the reconstructed data (the red dashed line). Our model captures 99.23% of the variance that could be captured by an optimal linear model with 100 output neurons.

References

Atick, J. J., & Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation*, *4*(2), 196–210.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3), 183–193.

Baldi, P., & Hornik, K. (1989). Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks*, *2*, 53–58.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 217–234). Cambridge, MA, USA: MIT Press.

Bartlett, M. S., Lades, M., & Sejnowski, T. J. (1998, January). Independent component representations for face recognition. *Proceedings of the SPIE Symposium on Electonic Imaging: Science and Technology; Conference on Human Vision and Electronic Imaging III*, *3299*.

Brady, N., & Field, D. J. (1995). What's constant in contrast constancy? the effects of scaling on the perceived contrast of bandpass patterns. *Vision Research*, *35*(6), 739–756.

Brady, N., & Field, D. J. (2000). local contrast in natural images: normalisation and coding efficiency. *Perception*, *29*(9), 1041–1055.

Buchsbaum, G., & Gottschalk, A. (1983). Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proceedings of the Royal Society B: Biological Sciences*, *220*(1218), 89–113.

Calkins, D. J., & Sterling, P. (1999, October). Evidence that circuits for spatial and color vision segregate at the first retinal synapse. *Neuron*, *24*, 313–321.

Carney, L. H. (1990). Sensitivities of cells in anteroventral cochlear nucleus of cat to spatiotemporal discharge patterns across primary afferents. *Journal of neurophysiology*, *64*(2), 437–456.

Caywood, M. S., Willmore, B., & Tolhurst, D. J. (2004, June). Independent components of color natural scenes resemble V1 neurons in their spatial and color tuning. *Journal of Neurophysiology*, *91*(6), 2859–2873.

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network: Computation in Neural Systems*, *12*(2), 199–213.

Chichilnisky, E. J., & Kalmar, R. S. (2002). Functional asymmetries in ON and OFF ganglion cells of primate retina. *Journal of Neuroscience*,

22(7), 2737–2747.

Cottrell, G. W., Munro, P., & Zipser, D. (1989). Image compression by back propagation: An example of extensional programming. In N. E. Sharkey (Ed.), *Models of cognition: A review of cognitive science* (pp. 208–240). Norwood, New Jersey: Ablex.

Dearworth Jr., J. R., & Granda, A. M. (2002). Multiplied functions unify shapes of ganglion-cell receptive fields in retina of turtle. *Journal of Vision*, *2*(3).

Derrico, J. B., & Buchsbaum, G. (1991). A computational model of spatiochromatic image coding in early vision. *Journal of Visual Communication and Image Representation*, *2*(1), 31–38.

Doi, E., Inui, T., Lee, T.-W., Wachtler, T., & Sejnowski, T. J. (2003). Spatiochromatic receptive field properties derived from information-theoretic analyses of cone mosaic responses to natural scenes. *Neural Computation*, *15*(2), 397–417.

Doi, E., & Lewicki, M. S. (2007). A theory of retinal population coding. In *Advances in neural information processing systems* (Vol. 19, pp. 353–360). Cambridge, MA, USA: MIT Press.

dos Santos, N. A., & Alencar, C. C. G. (2010, August). Early malnutrition diffusely affects children contrast sensitivity to sine-wave gratings of different spatial frequencies. *Nutritional Neuroscience*, *13*(4), 189–194.

Emmons, L. H., Whitney, B. M., & Ross, D. L. (n.d.). *Sounds of neotropical rainforest mammals [audio cd] (library of natural sounds, cornell laboratory of ornithology, ithaca, new york, 1997).*

Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology*, *187*(3), 517–552.

Fairchild, M. D. (2001). A revision of CIECAM97s for practical applications. *Color Research & Application*, *26*(6), 418–427.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of American, A*, *4*(12), 2379–2394.

Field, D. J., & Brady, N. (1997). Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research*, *37*(23), 3367–3383.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*, 1–22.

Graham, D. J., Chandler, D. M., & Field, D. J. (2006). Can the theory of

"whitening" explain the center-surround properties of retinal ganglion cell receptive fields? *Vision Research*, *46*(18), 2901–2913.

Griffin, G., Holub, A., & Perona, P. (2007). *Caltech-256 object category dataset* (Tech. Rep. No. 7694). California Institute of Technology. Retrieved from `http://authors.library.caltech.edu/7694`

Jolliffe, I. T. (2002). *principal component analysis*. Spring-Verlag.

Karklin, Y., & Lewicki, M. S. (2009). Emergence of complex cell properties by learning to generalize in natural scenes. *Nature*, *457*, 83–86.

Kiang, N. Y.-S., Watanabe, T., Thomas, E. C., & Clark, L. F. (1965). *Discharge patterns of single fibers in the cat's auditory nerve*. MIT Press Cambridge.

Lamel, L. F., Kassel, R. H., & Seneff, S. (1986). Speech database development: design and analysis of the acoustic-phonetic corpus. In *Proceedings of the DARPA speech recognition workshop* (pp. 100–109).

Lee, T.-W., Wachtler, T., & Sejnowski, T. J. (2002). Color opponency constitutes a sparse representation for the chromatic structure of natural scenes. *Vision Research*, *42*(17), 2095–2103.

Lewicki, M. S. (2002). Efficient coding of natural sounds. *Nature Neuroscience*, *5*(4), 356–363.

Lewicki, M. S., & Olshausen, B. A. (1999). A probabilistic framework for the adaptation and comparison of image codes. *Journal of the Optical Society of America A*, *16*(7), 1587–1601.

Lewicki, M. S., & Sejnowski, T. J. (2000). Learning overcomplete representations. *Neural Computation*, *12*(2), 337–365.

Mairal, J., Bach, F., Ponce, J., & Sapiro, G. (2010). Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, *11*, 10–60.

Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001, July). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the 8th international conference on computer vision* (Vol. 2, pp. 416–423).

Olmos, A. A., & Kingdom, F. A. A. (2004). *Mcgill calibrated colour image database*. `http://tabby.vision.mcgill.ca`.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607–609.

Olshausen, B. A., & O'Connor, K. N. (2002). A new window on sound. *Nature*, *5*(4), 292–294.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., & Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, *454*(7207), 995–999.

Recio-Spinoso, A., Temchin, A. N., van Dijk, P., Fan, Y.-H., & Ruggero, M. A. (2005). Wiener-kernel analysis of responses to noise of chinchilla auditory-nerve fibers. *Journal of neurophysiology*, *93*(6), 3615–3634.

Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, *439*(7079), 800–805.

Smith, T., & Guild, J. (1931). The CIE colorimetric standards and their use. *Transactions of the Optical Society*, *33*, 73–134.

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *216*(1205), 427–459.

Stokes, M., Anderson, M., Chandrasekar, S., & Motta, R. (1996). A standard default color space for the Internet–sRGB. *Microsoft and Hewlett-Packard Joint Report*.

Tailor, D. R., Finkel, L. H., & Buchsbaum, G. (2000). Color-opponent receptive fields derived from independent component analysis of natural images. *Vision Research*, *40*(19), 2671–2676.

van Hateren, J. H., & Ruderman, D. L. (1998). Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proceedings of the Royal Society B: Biological Sciences*, *265*(1412), 2315–2315.

van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society B: Biological Sciences*, *265*(1394), 359–366.

Vincent, B. T., & Baddeley, R. J. (2003). Synaptic energy efficiency in retinal processing. *Vision Research*, *43*(11), 1283–1290.

Vincent, B. T., Baddeley, R. J., Troscianko, T., & Gilchrist, I. D. (2005). Is the early visual system optimised to be energy efficient? *Network: Computation in Neural Systems, special issue on Sensory Coding and the Natural Environment*, *16*(2/3), 1283–1290.

Wachtler, T., Lee, T.-W., & Sejnowski, T. J. (2001). Chromatic structure of natural scenes. *Journal of the Optical Society of America A*, *18*(1),

65–77.

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of computational and graphical statistics*, *15*(2), 265–286.