Proceedings of the 2005 IEEE
Engineering in Medicine and Biology 27th Annual Conference
Shanghai, China, September 1-4, 2005

A Filter Based Encoding Model For Mouse Retinal Ganglion Cells

Q. Zhong*, P. O. Boykin†, A. Jacobs‡, V. P. Roychowdhury*, S. Nirenberg‡
* Department of Electrical Engineering, University of California, Los Angeles, CA, USA
† Department of Electrical and Computer Engineering, University of Florida, Gaineville, FL, USA
‡ Department of Neurobiology, University of California, Los Angeles, CA, USA

Abstract— **We adopt a system theoretic approach and explore the model of retinal ganglion cells as linear filters followed by a Maximum-Likelihood Bayesian predictor. We evaluate the model by using cross-validation, i.e., first the model parameters are estimated using a training set, and then the prediction error is computed (by comparing the stochastic rate predicted by the model with the rate code of the response) for a test set. As in system identification theory, we present spatially uniform stimuli to the retina, whose temporal intensity is drawn independently from a Gaussian distribution, and we simultaneously record the spike trains from multiple neurons. The optimal linear filter for each cell is obtained by maximizing the mutual information between the filtered stimulus values and the output of the cell (as measured in terms of a stochastic rate code). Our results show that the model presented in this paper performs well on the test set, and it outperforms the identity Bayesian model and the traditional Linear model. Moreover, in order to reduce the number of optimal filters needed for prediction, we cluster the cells based on the filters' shapes, and use the cluster consensus filters to predict the firing rates of all neurons in the same class. We obtain almost the same performance with these cluster filters. These results provide hope that filter-based retinal prosthetics might be an effective and feasible idea.**

I. INTRODUCTION

The quality and quantity of experimental data from biological neuronal systems is dramatically rising [1]. With more data, there is a need for algorithms which are guided by the data, rather than guided by human analysis of that data. Information theory [2], [3] then becomes a natural framework to study how neurons communicate information in the nervous system. It introduces the *mutual information* between two random variables S and R, denoted as $I(S;R)$, which is a measure of the statistical dependency. It is also the upper bound on how many bits about S (or R) we know on average. In spite of many advances, it is still not known precisely how information is represented by neural code [4], [5].

In this work, we model the retina ganglion cells of a mouse as linear filters using a stochastic rate code to encode the filter values. With the genetic algorithm, we find the filter which maximizes mutual information between the filtered stimulus and the number of spikes in a time frame. We simply describe our experimental data in section II. In section III, we specifically introduce our motivation and the optimal filters for information maximization. Followed in section IV, we evaluate the encoding performance for three encoders. We have also shown that using the cluster filters, we can improve the encoding performance as well.

II. DATA DESCRIPTION

Mice were placed in the dark overnight before the experiment then killed with 100% CO2. The eye was then removed. And under dim red light (low-pass filter with a 580 nm cut of; intensity, 0.65 $microwatts/cm^2$ or 250 rod-equivalent $photons/micron^2/sec$) the cornea, lens, and vitreous were removed and the retina was isolated from the pigment epithelium. A 2.5×2.5 mm piece was cut from the central retina and placed on a 64-electrode extracellular multielectrode array in a recording chamber where it was perfused continuously with oxygenated Ringer's solution throughout the experiment. The cells responses were then recorded simultaneously. The experiment was performed in accordance with UCLA animal research committee guidelines.

The data set we analyzed consists of 17 retinal ganglion cells from a mouse. The stimulus was the full field of a computer monitor with a 256-value linear intensity scale (0 is darkest, 255 is brightest), driven by a series of Gaussian deviating with a mean of 127 and a standard deviation of 60 every 100 ms for 1.5 hour. The firing rate ranges from 4.2 Hz to 77.4 Hz.

III. METHODOLOGY AND TECHNIQUES

A. Motivation

It takes time for a neuron to transfer the visual information, thus the neural response at the present frame might not be most informative about the present stimulus, but instead about an earlier stimulus. In addition, the response at one frame might be informative about stimulus at several frames. We then want to investigate this encoding time process using the information theory.

The stimulus is then represented as a vector $S = (s_1, s_2, \ldots, s_N)$, where s_i is the intensity of frame i. We binned the stimulus intensity values into 16 approximately equal sized bins to reduce the number of bits per frame to be at most 4. For each cell, we also binned the spike train in 100 ms bins (a stimulus frame) and counted the number of spikes in every bin. Then the response of the j^{th} cell is $R^j = (r_1^j, r_2^j, \ldots, r_N^j)$, where r_k^j is the number of spikes in the k^{th} bin for the j^{th} cell. Ignoring any temporal dependencies, we might consider the mutual information $I(S, R^j)$ created by pairing each (s_i, r_i^j) for all i. We could calculate

$$I(S;R^j) = H(S) + H(R^j) - H(S, R^j). \qquad (1)$$

We also introduced the $R_l^j = (r_{1+l}^j, r_{2+l}^j, \ldots, r_N^j)$, which has the length $N - l$. If $l \ll N$, then $H(R^j) \approx H(R_l^j)$. For each delay l, the mutual information $I(S; R_l^j)$ could be computed. The encoding efficiency [6] with delay l is defined as

$$\eta_l = I(S; R_l^j)/H(R_l^j). \tag{2}$$

Fig. 1 shows the encoding efficiency with different delays for every neuron. Note that no cell has a maximum of η_l at $l = 0$. Instead, all of them reach the maximum at $l = 1$ or $l = 2$. Thus all the cells in this study are most informative about the stimulus $100 - 300$ ms in the past. Since that our stimulus was drawn independently, the mutual information being spread over a wide time window is due only to neural processing instead of the statistics of the stimulus.

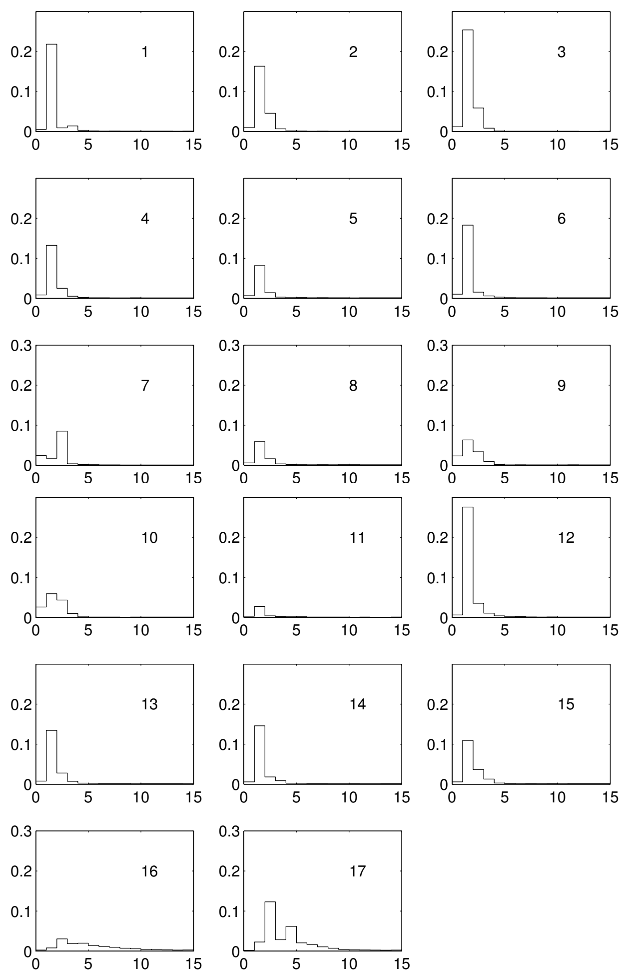

Fig. 1. Encoding Efficiency with Different Delays. Every subplot represents a neuron. The x-axis corresponds to the number of delay frames.

B. *Linear filter for mutual information maximization*

In our setting, a general model for the cell as a rate-encoder was to assume that the probability $p_j(r_i|s_i, s_{i-1}, \ldots s_{i-k})$ for each cell j exists, where k captures how much past dependency there is in each cell. In Fig. 1 we see that many cells have non-zero information about stimuli 500 ms or more into the past. This indicates that several frames of stimulus are processed jointly to generate its subsequent response. Thus we would need to use at least $5 - 10$ frames of past data to predict the response at a given time. Given that we had access to only $54,000$ frames, we had a woefully insufficient amount of data to build such a model.

Thus instead, we assumed that the cell depends on a convolution, or filter, of the stimulus

$$s_w(i) = \sum_{j=0}^{k} w_j s_{i-j}, \tag{3}$$

and instead of the cell being defined by a table given by $p_j(r_i|s_i, s_{i-1}, \ldots s_{i-k})$, we considered the case where $p_j(r_i|s_w(i))$ exists.

How could we use the data to select the best w? Our method was to find the w_{opt}^j such that

$$I(S_w; R^j) \leq I(S_{w_{opt}^j}; R^j). \tag{4}$$

Equivalently, we wanted the w_{opt}^j to maximize the mutual information between the filtered stimulus and the response.

The mutual information is not affected by any constant scaling of w. To make sure all filters w are comparable we normalized $\sum_i | w_i |= 1$, which requires that $-1 \leq w_i \leq 1$. Additionally, we made sure that the filtered value had a positive correlation with the response by setting the overall sign of w. This differentiates the cells positively correlated with the stimulus from those negatively correlated with the stimulus. Thus we have a model that includes potentially long histories to the cell with only linearly many parameters to fit, rather than exponentially many.

We knew of no analytic solution to finding w_{opt}^j such that (4) holds. Instead computational optimization techniques must be used. In our analysis we chose k to be 14, which means that we considered the 15 frames of stimulus (1.5 s preceding the response).

We focused on solutions given by a Genetic algorithm [7]–[10]. Beginning with 80 parent individuals generated from the uniform distribution over $[-1,\ 1]$, after 400 generations evolution we chose the filter with the highest mutual information as the approximate optimal solution[1].

Shown in Fig. 2, we see that the filter displays its positive or negative peak at the 2^{nd} or 3^{rd} frame. Thus two different measures, encoding efficiency (Fig. 1) and the optimal filter (Fig. 2) both suggest the same processing time, 100 ms to 300 ms, called the best offset, for the mouse retinal ganglion cells to convey information.

Compared with $I(S; R_l^j)$, the mutual information between the stimulus and the response with the optimal offset, $I(S_{w_{opt}^j}; R^j)$ increases around $50\% - 300\%$ after the filtering process. It's value of clarifying whether the increase of

[1]We repeated this analysis 10 times and the Euclidean distance between different simulation results is 0.029 on the average of all 17 neurons, while the distances among these 17 neurons range from 0.0353 to 1.0826. We also used the method of simulated annealing [11], [12] and found similar results, but the computation required much longer to find the optimum.

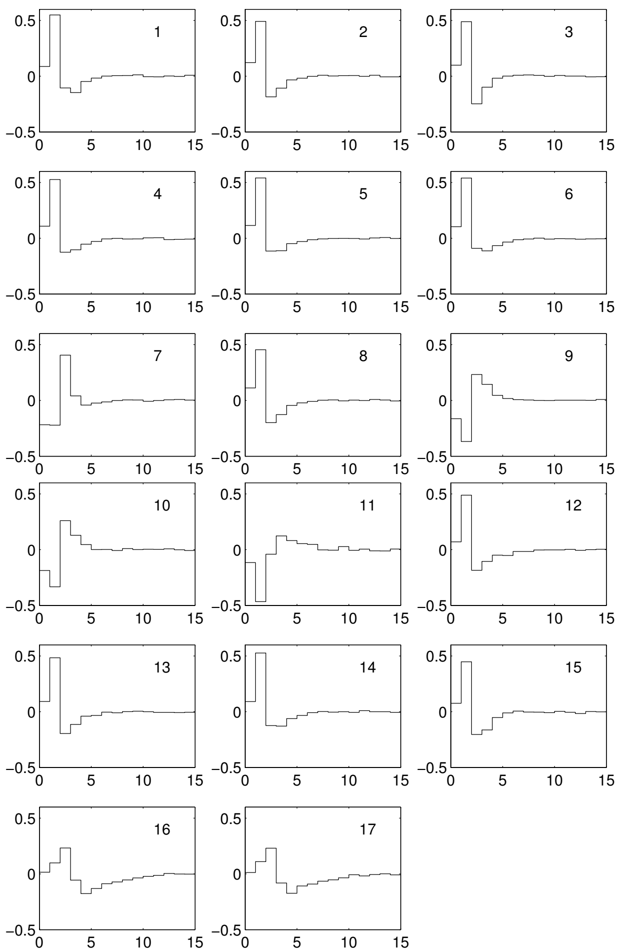

Fig. 2. Optimal linear filters of retinal ganglion cells. Every subplot represents a neuron. The x-axis corresponds to the number of delay frames, so the filter is displayed in the order of $w_0, w_1, \cdots, w_{14}$. We have normalized the vector such that $\sum_i | w_i |= 1$.

the mutual information is due to the $H(S_{w^j_{opt}})$ increase or not. However, $H(S_{w^j_{opt}})$ is 86.6% of $H(s)$ averaged on all these neurons. Thus the filtering process emphasizes the information that the neurons carry.

IV. RESULTS

A. Encoding performance

We evaluate three rate encoding models by using them to predict the firing rate here. To make our results robust, we did the cross validation. The whole data set was then split into 2 halves.

- Identity Bayesian Model (IB): We only used the stimulus at some fixed point in the past to predict the number of spikes in the current bin. And usually we are only interested in the best case where IB is with its optimal offset, an offset of one or two frames in the past. We could also choose the set of frames which have the highest encoding efficiency to predict (IB(i) means that we choose i best frames), but our data size limited our choice of 2 best frames at most.

TABLE I
THE AVERAGE ENCODING PERFORMANCE FOR THREE MODELS.

	IB(1)	IB(2)	LR	FB
ER	0.53	0.41	0.37	0.32
CORR	0.50	0.62	0.59	0.73

- Linear Model (LR): Inspired by the linear decoding method [6], we could obtain a linear filter by minimizing the mean square error between the linear filtered stimulus value and the neuron's firing rate.
- Filter Based Bayesian Model (FB): For each cell, we applied its optimal filter which was obtained by maximizing the mutual information $I(S_w; R)$. So instead of the mean square error used in LR, the cost function in FB is mutual information. We obtained $\tilde{r}$ by using the maximum likelihood Bayesian algorithm for both IB and FB.

The average encoding performance[2] of the four encoding models is given in Table I, where the normalized mean square error is denoted as ER and the correlation coefficient is denoted as CORR. The average ER is around 60% as large for FB as for IB(1), and the average CORR has a 46% increase. Using two best frames, IB(2) does a better job than IB(1), which emphasizes the importance of combining several stimulus frames for prediction, and it provides a comparable performance to LR. We realize that LR provides a comparable ER to FB, mostly because its cost function is to minimize the mean square error, but it's still outperformed by FB by giving a much smaller CORR. Thus in general, we can order the performance of these four models as $FB > LR > IB$.

B. Encoding with cluster filters

We also considered of reducing the number of filters needed in prediction. For this aim, we clustered the neurons based on the shape of their optimal filters via the agglomerative hierarchical clustering technique [13] at first. Every neuron's filter is viewed as a vector, then the distance between two filters is defined to be either the CityBlock distance d_1 or the Euclidean distance d_2, and the distance between two clusters to be either the average distance D_{ave}, or the maximum distance D_{max}. The hierarchical clustering algorithm started from every neuron as a single cluster, and we initially merged two clusters i and j that had the minimal distance value. We then reduced the number of clusters by one. We re-calculated the distances between clusters and iterated the above step until we got a single cluster. The result of clustering of the neurons with Euclidean filter distance and average cluster distance is shown in Fig. 3a as a binary

[2] We have excluded the outlier cells 11, 14 and 16 for a fair comparison of the four encoding models. Both IB and FB don't work for these cells.

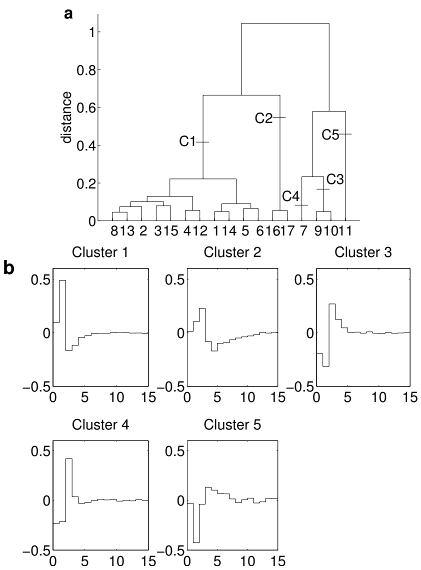

Fig. 3. We have clustered the neurons according to optimal filters derived from the $1st$ half of the data. a: Dendrogram of the neurons' clustering. The height of a merge represents the distance between two clusters. The 5 important clusters are highlighted. b: The 5 cluster filters. Note cluster 4 and cluster 5 are single clusters.

dendrogram, in which we found 5 important clusters in total[3]. With this dendrogram, we define the cluster filter of each cluster to be

$$w_C = \frac{1}{|C|}\sum_{k \in C} w_k, \tag{5}$$

as shown in Fig. 3b. Now we could use those 5 most important cluster filters to predict. We have obtained similar performance with these cluster filters: the average ER is 0.33 and the CORR is 0.72.

V. SUMMARY

The filter based Bayesian model is a general model to describe the encoding behavior of the cell. Our model of the neurons as rate-encoders of the optimal filtered stimulus can do a good job in predicting the neurons' firing rate. Compared with IB(1), it cuts the error rate by around 40% (0.32 versus 0.53), and raises the correlation coefficient by 46% (0.73 versus 0.50). In addition, it outperforms LR, whose performance is comparable to that of IB(2), by giving a less error rate (0.32 versus 0.37) and a much larger correlation coefficient (0.73 versus 0.59). Moreover, we could reduce the number of optimal filters we need for encoding by doing the hierarchical clustering. We could do an almost same job with these cluster filters as FB (error rate: 0.33 versus 0.32; correlation coefficient: 0.72 versus 0.73).

There are several areas where our filter based model can be generalized. First, instead of using a spatially uniform stimulus, we may apply our technique to a non-spatially uniform stimulus. In this case, the filter size would grow dramatically (to cover both space and time). A second approach would be to extend our linear filter to a class of non-linear filters and do the same analysis. Finally, we can have a number of filters assigned to each cell and find the optimal set of filters whose outputs have the most mutual information with the response.

We should also note that the stimulus is a very challenging one to decode: white gaussian noise at 10 Hz. This has a broader spectrum than one would expect an animal to encounter in the wild. It will be interesting to apply this technique to a more slowly changing stimulus. Deconvolution techniques may allow us to reconstruct the stimulus very precisely from the several classes of cells, if that stimulus is not too high in frequency.

[3]The clustering technique applied to the filters is very robust: *all four choices gave rise to the same* 5 *most significant clusters.* In fact, we also found evidence that the filters as well as the clustering structure are stationary, and the clustering structure is not dependent on firing rate.

ACKNOWLEDGMENT

This work was supported in part by the NSF grant ITR 0326605.

REFERENCES

[1] M. Meister, J. Pine, and D.A. Baylor, "Multi–neuronal signals from the retina: acquisition and analysis," *J. Neurosci. Methods.*, vol. 51, pp. 95–106, 1994.
[2] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Illinois: Univ. Illinois Press, 1949.
[3] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.
[4] F. Rieke, D. Warland, R.R. Steveninck, and W. Bialek, *Spikes:Exploring the Neural Code.* Cambridge, MA: MIT press, 1997.
[5] M. Meister and M.J. Berry, "The neural code of the retina," *Neuron*, vol. 22, pp. 435–450, 1999.
[6] D.K. Warland, P. Reinagel, and M. Meister, "Decoding visual information from a population of retinal ganglion cells," *J. Neurophsiol.*, vol. 78(5), pp. 2336–2350, 1997.
[7] J. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor: The university of Michigan press, 1975.
[8] D.E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley Publishing Company, 1989.
[9] A.J. Chipperfield, P.J. Fleming, and C.M. Fonseca, "Genetic algorithm tools for control systems engineering," Plymouth Engineering Design Centre, September 1994, pp. 128–133, presented at Adaptive Computing in Engineering Design and Control.
[10] A.J. Chipperfield and P.J. Fleming, "The MATLAB genetic algorithm toolbox," January 1995, from IEE Colloquium on Applied Control Techniques Using MATLAB, Digest NO. 1995/014.
[11] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21(6), pp. 1087–1092, 1953.
[12] S. Kirkpatrick and C.D. Gelatt and M.P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220,4598, pp. 671–680, 1983.
[13] J.H. Ward, "Hierarchical grouping to optimize an objective function," *J. Am. Stat. Assoc.*, vol. 58, pp. 236–244, 1963.