Adaptive contrast gain control and information maximization [1]

Yuguo Yu[*] and Tai Sing Lee[*,†]

Center for the Neural Basis of Cognition [*] *& Computer Science Department* [*,†]
Carnegie Mellon University, Pittsburgh PA 15213, U.S.A.

Abstract

Contrast gain control is an important mechanism underlying the visual system's adaptation to contrast of luminance in varying visual environments. Our previous work showed that the threshold and saturation determine the preferred contrast sensitivity as well as the maximum information coding capacity of the neuronal model. In this report, we investigated the design principles underlying adaptation behavior in contrast gain control by an adaptive linear-nonlinear model. We found that an adaptive rescaling mechanism predicted by information transmission maximization can explain a variety of observed contrast gain control phenomena in neurophysiological experiments, including the divisive input-output relations, and the inverse power law relation between response gain and input contrast. Our results suggest that contrast gain control in visual systems might be designed for information maximization.

Key words: Adaptation, Contrast gain control, Information maximization

1 Introduction

The visual systems exhibit great flexibility in adapting their input-output functions to the variance or the contrast [1-4] of light signal in the visual environment. The amplitude gain of the transfer function of visual neurons was found to decrease with input variance [1], and even displays an inverse power law relation [4] (see Fig.2). In addition, the contrast response function in visual cortical neurons was found to shift along the log contrast axis to adaptively match the prevailing input signals [2]. These contrast adaptation phenomena have been widely observed in the retina [1], striate [2] and extrastriate visual

cortex [5] of mammals, and fly H1 neurons [6,7], suggesting the existence of an contrast gain control mechanism. However, the biophysical factors and computational rules governing its operation remain elusive. How to understand these phenomena? Why visual neurons behavior in this way?

We have shown that classical neuronal model with nonlinearities of threshold and saturation can display a apparent adaptation in the recovered transfer function to the variance of the input signal [8,9]. However, we found that the mutual information in those system, as well as the kernel gain, exhibit a bell-shape as a function of the input variance [2]. There is only one optimal variance can induce the maximal information transmission of the system. That means nonlinear dynamics due to static nonlinearity alone cannot maintain a neuron's information transmission rate at the highest level with the variation in signal variance. Here, we explored theoretically a possible extra adaptation mechanism that might restore information maximization for any input variance in neurons. We introduce an additional gain rescaling to the static model, while keeping all other parameters (thresholds, saturation levels, membrane time constants) constant. We found it sufficient to produce information maximization that is consistent with some neurophysiological findings [6,7]
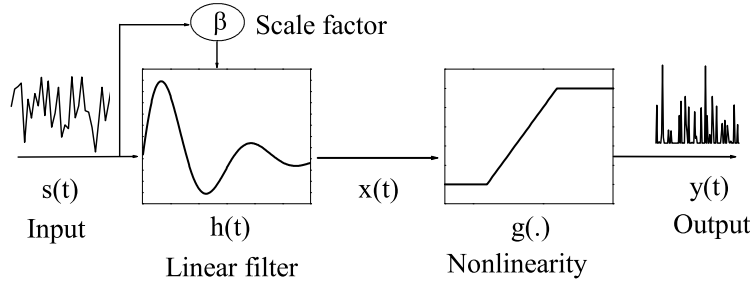
2   Model and Results

Fig. 1. The adaptive LN model consists of a linear filter $h(t)$ followed by a nonlinearity $g(.)$. The amplitude of the linear filter $h(t)$ is controlled by β, which acts as an adaptive mechanism. $x(t)$ is the convolution of the input signal $s(t)$ and the filter $h(t)$. The nonlinearity $g(.)$ operates on $x(t)$ to generate the output $y(t)$.

We use the similar standard linear-nonlinear (LN) cascade model as in Ref.[9] except that we introduce an additional scaling factor $\beta(\sigma)$ see Fig.1. β scales the amplitude of the linear kernel according to the variance of the input signal as an adaptive mechanism. Linear kernel is given by $h(t) = \beta(\sigma) * A * sin(\pi t/\tau_a)exp(-t/\tau_b)$ with $A = 1$, $\tau_a = 80$ ms and $\tau_b = 100$ ms. The output of linear kernel $x(t)$ is given by $x(t) = \int_0^{+\infty} h(\tau)s(t - \tau)d\tau$. The

nonlinearity $g(.)$ is specified by

$$
g(x) = \begin{cases} 0, & \text{if } x < \theta, \\ x - \theta, & \text{if } \theta \le x < \eta, \\ \eta - \theta, & \text{if } x \ge \eta. \end{cases} \tag{1}
$$

where θ is threshold and η is saturation level. Without loss of generality, we use a Gaussian white noise stimulus $s(t)$ with zero mean and SD σ as the input signal. Its probability density function (PDF) is given by $p(s) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{s^2}{2\sigma^2}}$. The linear response $x(t)$ also has a Gaussian distribution with PDF $p(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{x^2}{2\sigma_x^2}}$, where σ_x is given by $\sigma_x^2 = <x^2(t)> = \sigma^2 \int_0^{+\infty} h^2(\tau)d\tau$, where $< ... >$ denotes time average. β is determined computationally by maximizing the mutual information between $s(t)$ and $y(t)$ with respect to β, for signals of a given σ. This is the adaptive LN model.

We have shown the amplitude gain of the recovered linear kernel of the static model is a function of the input signal variance σ^2 [7]. The ratio α between the amplitude gain of the recovered linear kernel $h'(t)$ and the original linear kernel $h(t)$ is given by

$$
\alpha = \frac{\int_\theta^\eta x(x - \theta)p(x)dx + (\eta - \theta) \int_\eta^{+\infty} x p(x)dx}{\sigma^2 \int_0^{+\infty} h^2(\tau)d\tau}. \tag{2}
$$

That is, for an input signal with Gaussian white distribution, we have $h'(t) = \alpha.h(t)$, where gain factor α quantifies how the performance of linear function $h(t)$ is affected by the static nonlinearity (threshold θ, saturation η) and stimulus standard deviation σ. Mutual information [10] quantifies the ability of a system or a communication channel to convey information and is given by

$$
\begin{aligned} I_m &= H(y) - H(y \mid s) \\ &= -\sum_y P(y)log_2 P(y) + \sum_{s,y} P(s)P(y \mid s)log_2 P(y \mid s), \end{aligned} \tag{3}
$$

where $H(y)$ is the total entropy of the response, $H(y \mid s)$ is noise entropy, accounting for the variability in the response that is not due to variations in the stimulus, but comes from other noise sources. For simplicity, we consider the noiseless case, where $H(y \mid s) = 0$. In this case, the mutual information is simply equal to the output entropy $I_m = H(y)$.

We are interested in computing the entropy of the output y. The distribution

function of y (also refereed to as the cumulative density function) is

$$D(y) = P[Y \leq y] = \begin{cases} 0, & \text{if } y < 0, \\ \int_{-\infty}^{y+\theta} p(x)dx, & \text{if } 0 \leq y < \eta - \theta, \\ 1, & \text{if } y \geq \eta - \theta. \end{cases} \quad (4)$$

In order to compute the entropy of y, we approximate y by a discrete random variable $\hat{y}$, which we define by $P[\hat{y} = i\,\Delta y] = P\big[y \in [(i-1)\,\Delta y, i\,\Delta y]\big]$, where Δy is the quantization resolution of y, and i is an integer from 0 to $\lceil(\eta - \theta)/(\Delta y)\rceil$. Note that $P[y \in (a, b]\,]$ can be computed analytically as $D(b) - D(a)$. The mutual information, which is equal to entropy for the noiseless case, is then computed as in Eq.[3]. It can be shown that changes in Δy affect the entropy of $\hat{y}$ only by an additive constant [11]. Since we are only interested in comparing the values of mutual information I_m as we change σ, θ, and η, our choice of Δy does not affect our results. Here, we set $\Delta y = 1$.

Figure 2b (blue line) shows mutual information I_m varies with σ, exhibiting a tuning curve, with a global maximum at a particular σ. This maximum is determined by the threshold and the saturation levels [8,9] The lower (or higher) is the threshold (saturation), the larger the maximum of information can be transmitted (see Fig.2c). Can rescaling the gain of the linear kernel β restores the information transmission to its maximum possible level at different σ's? This is desirable because the capacity of the system will then be fully utilized regardless of the input variance. We found that rescaling β alone can indeed achieve this purpose. Fig. 2a shows that for each σ, mutual information varies with the scaling factor β and that there always exists an optimal β to restore the information transmission rate of the system to the maximum level which is constrained only by the threshold and the saturation level (Fig 2c). This gain rescaling allows the system's information transmission rate to be maintained at optimal regardless of the input variance (Fig 2b). The optimal β decreases monotonously with an increase in σ. Interestingly, in the log-log plot, β and σ displays a inverse power law relationship, the fitting slope is -1 (Fig. 2d).

The response gain γ, i.e., the amplitude of the recovered kernel, now equal to $\alpha * \beta$ (see Eq.[2]), also displays a scale invariant power law relationship with a scale -1 with σ (Fig.3a). Noted that, without gain rescaling, i.e., $\beta = 1$ is fixed, the gain γ varies with σ, displaying a bell-like tuning curve, which emerges simply from static-nonlinearity. Gain rescaling is crucial for maintaining mutual information constant with variation in σ. There is experimental evidence supporting this inverse-power law between recovered kernel amplitude gain and the signal contrast in the neurophysiological literature. Fig.3b is adapted from Fig.5a in Ref.[4], which shows that response gain of cat simple cell varying with the contrast of monocular input also in power law relationship in the log-contrast plot. A contrast gain slop of -1 indicates that the
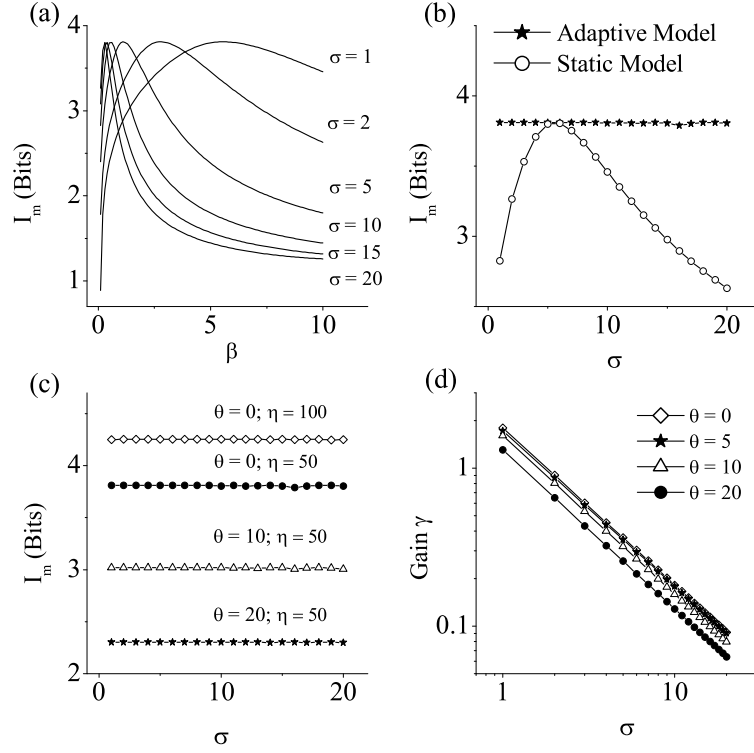
Fig. 2. (a) For an adaptive LN model with threshold $\theta = 0$ and saturation $\eta = 50$, mutual information I as a function of rescaling factor β for signals of different σ's. (b) The mutual information I as a function of σ for the adaptive LN model and the static LN model. (c) Mutual information I as a function of σ for various model parameters with β chosen to maximize I for each σ. (d) Optimal β as a function of σ for four given values of threshold in a log-log plot exhibits an inverse power-law.

gain control system is 100% effective, whereas a slope of 0 indicates that the system is completely ineffective. The cells in Fig. 3b showed a slope around -0.75 suggesting the neurons might be trying to re-scale gain adaptively in order maximize information transmission.

3 Discussion

The power law relationship observed experimentally and the theoretical findings of this paper suggests that there might exists an intrinsic adaptation mechanism in single neurons or network to maximize the information encoding. Such an adaptation, which is independent of static nonlinearity, involves mainly changing of the amplitude of the linear transfer function. Biophysically, a single neuron itself is a circuit, whose transfer function is controlled by the ionic conductances and capacitances. Recent experiments showed that the activation of Na^+-activated and Ca^{2+}-activated K^+ currents inside single
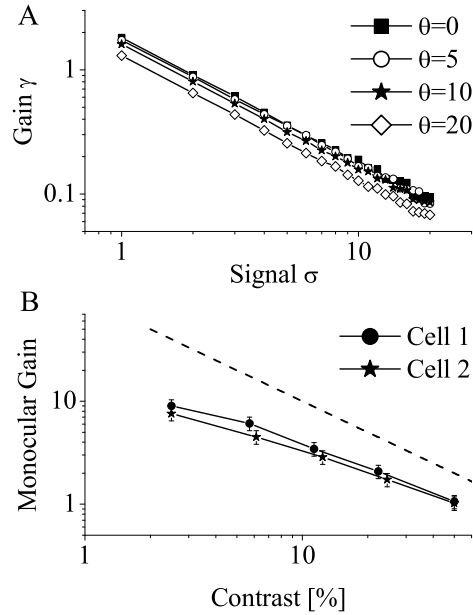
Fig. 3. Gain γ as a function of σ for four given values of threshold with β optimized for information transmission. (b) Monocular gain in two cat simple cells as a function of stimulus contrast (from 2.5% to 50%) in the log-log axes; for comparison, a line with slope of -1 is plotted in dashed line. Adapted from Fig 5. of Ref.[4].

cells [12,13] or the adjustment of voltage dependent conductance might play an important role in contrast adaptation [14]. These mechanisms might serve to control the gain of the linear kernel part of the LN cascade.

In summary, the contrast adaptation phenomena observed in neurophysiological experiments can potentially be factored into a component due simply to the static nonlinearities of the neurons [1-2] and another component due to an adaptive rescaling of the gain in the linear transfer function in the LN cascade for the purpose of information maximization.

References

[1] R.M. Shapley and J.D. Victor, The contrast gain control of the cat retina. *Vision Res.* 19 (1979) 431-434.

[2] I. Ohzawa, G. Sclar, and R.D. Freeman, Contrast gain control in the cat's visual system. *J. Neurophysiol.* 54 (1985) 651-667.

[3] S.M. Smirnakis, M.J. Berry, D.K. Warland, W. Bialek, and M. Meister, Adaptation of retinal processing to image contrast and spatial scale. *Nature* 386 (1997) 69-73.

[4] A.M. Truchard, I. Ohzawa, and R.D. Freeman, Contrast gain control in the visual cortex: monocular versus binocular mechanisms. *J. Neurosci.* 21 (2000) 287-299.

[5] A. Kohn and J.A. Movshon, Neuronal adaptation to visual motion in area MT

of the macaque. *Neuron* 39 (2003) 681-691.

[6] N. Brenner, W. Bialek, and R. de Ruyter van Steveninck, Adaptive rescaling maximizes information transmission, *Neuron*, 26 (2000) 695-702.

[7] A.L. Fairhall, G.D. Lewen, W. Bialek, and R. de Ruyter van Steveninck, Efficiency and ambiguity in an adaptive neural code. *Nature* 412 (2001) 787-792.

[8] Y.G. Yu and T.S. Lee, Dynamical mechanisms underlying contrast gain control in single neurons *Phys. Rev. E* 68 (2003) 1901-1907.

[9] Y.G. Yu, B. Potetz and T.S. Lee, The role of static nonlinearity in gain tuning and information transmission. *Vision Research* Accepted, (2004).

[10] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (Univ. of Illinois Press, Ur-bana, IL, 1949).

[11] P. Dayan and L. F. Abbott, *Theoretical Neuroscience* (MIT Press, Cambridge, 2001), Chap. 4, pp.129.

[12] K.J. Kim and F. Rieke, Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells. *J. Neurosci.* 21 (2001) 287-299.

[13] M.V. Sanchez-Vives, L.G. Nowak, and D.A. McCormick, Cellular mechanisms of long lasting adaptation in visual cortical neurons in vitro. *J. Neurosci.* 20 (2000) 4286 C4299.

[14] M. Stemmler and C. Koch, How voltage-dependent conductances can adapt to maximize the information encoded by neuronal firing rate. *Nature Neurosci.* 2 (1999) 521-527.