

A quantitative model relating visual neuronal activity to psychophysical thresholds

Laurent Itti, Christof Koch and Jochen Braun ^{*}

Abstract – We investigate how a simple, physiologically motivated three-stage neuronal model can establish a quantitative relationship between activities in small populations of simulated early visual neurons and human psychophysical thresholds. The model consists of: First, a bank of linear filters tuned for orientation and spatial period; second, non-linear interactions between filters; and, third, a statistically efficient decision stage. The model quantitatively reproduces human thresholds for five classical pattern discrimination tasks, using a unique set of automatically determined parameters. The resulting model components are all plausible in terms of putative neuronal correlates.

1 INTRODUCTION

We propose a framework which tries to draw a consistent connection between neural responses and psychophysical thresholds. To this end, we have developed a simple computational model, many components of which are based on what we consider is the emerging consensus from both the psychophysical and physiological modeling literature [9,10,1,5,11]. Success in reproducing a

^{*} California Institute of Technology, Computation and Neural Systems Program, Pasadena, California 91125. {itti,koch,achim}@klab.caltech.edu

wide range of human thresholds with such a model would provide evidence for the hypothesis that psychophysical performance for simple pattern discrimination tasks mainly reflects early neuronal processing.

All the parameters in the model (11 total) are left freely adjustable. Five classical psychophysical experiments [9] – which investigate largely unrelated tasks such as increment contrast, orientation and spatial period discriminations – are then used to automatically constrain all model parameters. This approach has two goals: First, it will determine which of the model characteristics are necessary and sufficient to explain the psychophysical data. For example, is it necessary to assume non-linear processing, or will linear interactions between filters suffice? Second, it will allow us to study whether the model components inferred from human psychophysics are plausible in neuronal terms, such that a close correspondence between early visual neuronal processing and psychophysical thresholds can be established. For example, are the tuning bandwidths of the filters derived from psychophysical data in agreement with those measured electrophysiologically?

2 MODEL

The general architecture of the model is presented in Figure 1. In the first stage, a bank of Gabor-like linear filters analyzes a fixed location of the visual scene. Thus, the equivalent of one complete set of orientation columns in V1 is modeled (60 filters total, tuned for 12 orientations $\theta \in \Theta$ spanning 180 deg and 5 spatial periods $\lambda \in \Lambda$ spanning two octaves, and all with overlapping receptive

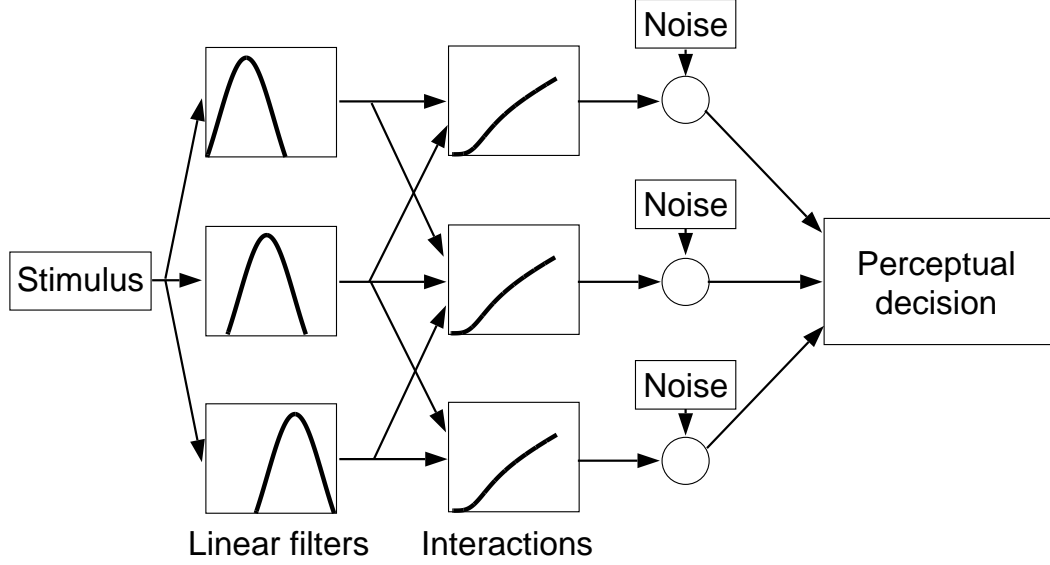


Fig. 1. General architecture of the model.

fields). Filters have Gaussian tuning for orientation and log spatial period, such that the linear response $E_{\lambda,\theta}$ of a unit tuned to (λ, θ) to a stimulus of contrast C_S , period λ_S and orientation θ_S is given by:

$$E_{\lambda,\theta}(C_S, \lambda_S, \theta_S) = C_S A \exp \left(-\frac{(\log(\lambda_S) - \log(\lambda))^2}{2\sigma_\lambda^2} - \frac{(\theta_S - \theta)^2}{2\sigma_\theta^2} \right) + \epsilon$$

where A is a constant gain coefficient, and ϵ a small constant to account for background activity.

In the second stage, filters non-linearly interact as follows: (1) Each unit receives non-linear self-excitation, and (2) each unit receives non-linear divisive inhibition from a pool of similarly-tuned units: With $E_{\lambda,\theta}$ given above for a unit tuned to (λ, θ) , the pooled response is given by:

$$R_{\lambda,\theta} = \frac{(E_{\lambda,\theta})^\gamma}{(S)^\delta + \sum_{(\lambda',\theta') \in \Lambda \times \Theta} W_{\lambda,\theta}(\lambda', \theta') (E_{\lambda',\theta'})^\delta} + \eta$$

where

$$W_{\lambda,\theta}(\lambda', \theta') = \exp\left(-\frac{(\log(\lambda') - \log(\lambda))^2}{2\Sigma_\lambda^2} - \frac{(\theta' - \theta)^2}{2\Sigma_\theta^2}\right)$$

is a 2D Gaussian weighting function centered around (λ, θ) , and η a positive constant to account for background activity in the pooling stage.

This stage is inspired from Heeger’s popular model of gain control in cat V1 [3,10,1]. Our formulation, in which none of the parameters is given a particular value, however allows for multiple outcomes, to be determined by fitting the model to our psychophysical data: A sigmoidal ($S > 0, \gamma > \delta$) as well as simple power-law ($S = 0, \gamma > \delta$) or even linear ($S = 0, \gamma = \delta + 1$) response characteristic could emerge, the responses could be saturating ($\gamma = \delta$) or not ($\gamma > \delta$), and the inhibitory pool size $(\Sigma_\lambda, \Sigma_\theta)$ could be broad or narrow. Because striate neurons are noisy, physiological noise is assumed in the model at the outputs of the second stage. The noise level is chosen close to what is typically observed in cortical pyramidal cells, i.e., with variance equal to the mean taken to some power $\alpha \approx 1$ [8] determined by fitting.

Because the decision stage – which quantitatively relates the activity in the population of noisy units in the second stage to behavioral discrimination performance – is not fully characterized in humans, we are not in a position to model it in any detail. Instead, we trained our subjects (for 2-3 hours on each task), and assume that they perform close to an “optimal detector”. Such optimal detector may be characterized in a theoretical manner, using the framework of Statistical Estimation Theory (see ref. [4]

for theoretical details). Let’s assume that a brain mechanism exists, which, for a given stimulus presentation, builds an internal estimate of some stimulus attribute ζ (e.g., contrast, orientation, period). The central assumption of our decision stage is that this brain mechanism will perform close to an “unbiased efficient statistic T ”, which is the best possible estimator of ζ given the noisy population response from the second stage. The accuracy with which T estimates ζ can be computed formally. Simply put, this means that, from the first two stages of the model alone, we have a means of computing the best possible estimation performance for ζ , and consequently [2], the best possible discrimination performance between two stimuli with parameters ζ_1 and ζ_2 . T is readily implementable as a neural network [6].

3 RESULTS and CONCLUSION

The 11 model parameters (the gain A of the linear stage, both tuning widths $\sigma_\lambda, \sigma_\theta$, both pooling widths $\Sigma_\lambda, \Sigma_\theta$, excitatory and inhibitory exponents γ, δ , background activities ϵ, η of the linear and of the pooling stage, activity-independent inhibition S , and noise exponent α) were automatically adjusted – using a 11-D downhill simplex [7] with simulated annealing overhead – such as to find the best fit of the model to the data (in the root-mean-square sense).

Remarkably, the model was able to simultaneously reproduce all the data, using a unique set of parameters (Figure 2). All parameters were constrained by the data and converged to reasonable values: The neuronal response as a function of contrast had the classical non-saturating sigmoidal shape; the orienta-

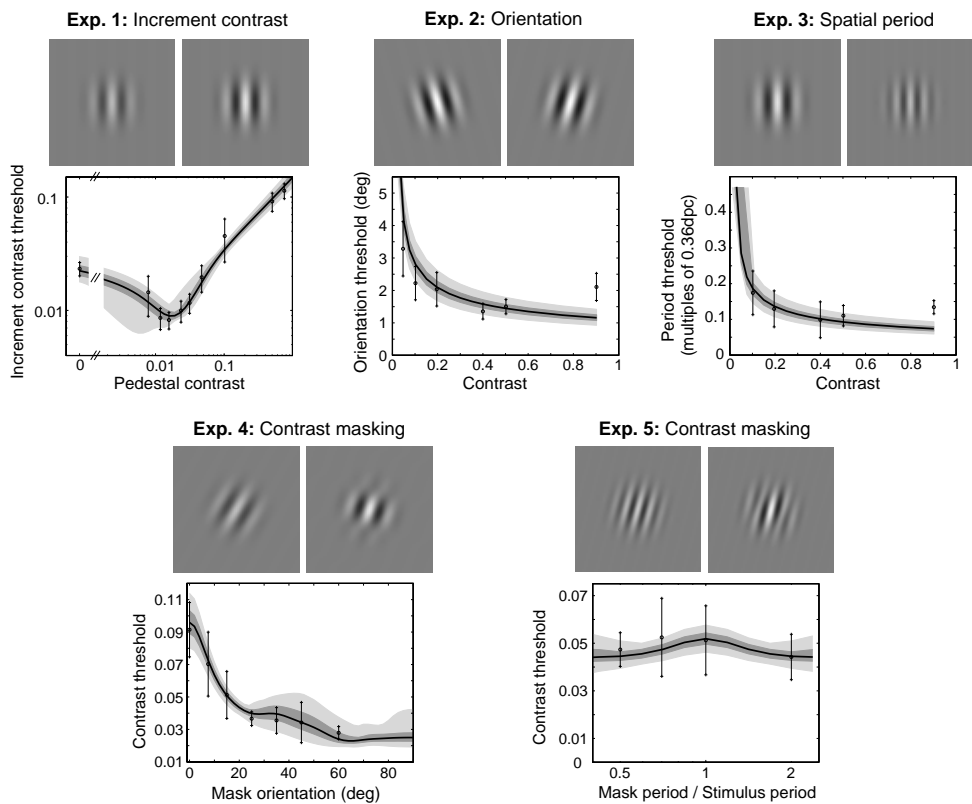


Fig. 2. Results of the automatic fitting of model parameters to human psychophysical data (two-alternative forced-choice paradigm, 2AFC) for five classical pattern discrimination tasks (one subject). For each experiment, both stimulus alternatives are shown (e.g., in Exp. 1, observers discriminate between a Gabor patch with fixed “pedestal” contrast and a second patch with slightly higher “pedestal+increment” contrast. Threshold is the value of the increment contrast yielding 75% correct discrimination). In Exps. 1–3, stimulus period is 2.8 cycles-per-degree (cpd). In Exp. 4 the mask is a Gabor patch with 50% contrast, 2.8 cpd period and variable orientation, and in Exp. 5, it has 50% contrast, variable period and 15 deg from vertical orientation. The shaded areas represent envelopes of all of the model predictions, when the model parameters were allowed to vary around their best-fit point such as to yield fitting errors lesser than 110% (narrow envelopes) or 150% (wider envelopes) the best-fit residual error. The regions of the datasets in which the envelopes are narrower around the best-fit curve more strongly constrain the model.

tion tuning full-width at half-maximum (FWHM) was 37deg for the linear units, yielding an FWHM of 24deg after pooling; spatial period tuning was 1.15 octaves before and 0.7 octaves after pooling; a narrow inhibitory pool was found for orientation

(FWHM=29deg), probably because no cross-orientation inhibition was observed in our data (Exp. 4), and a broad pool was found for spatial period (FWHM=5 octaves); finally, the noise level was slightly supra-Poisson (variance=mean^{1.1}).

Our success in reproducing a broad range of apparently unrelated human thresholds reinforces the idea that these thresholds result mainly from the earliest stages of visual processing. While our model was general enough to yield many functionally different outcomes, we found that one outcome – in good agreement with previous psychophysical and electrophysiological studies – emerged from an automatic best-fit of the model to our psychophysical data. In addition, our statistically efficient decision stage appeared as a particularly well suited tool for quantitatively relating neuronal population responses to psychophysical thresholds, using exactly the same theoretical assumptions for a potentially unlimited number of tasks.

References

- [1] J.M. Foley. Human luminance pattern-vision mechanisms: masking experiments require a new model. *J. Opt. Soc. A.* 1994;11(6):1710-9
- [2] D.M. Green and J.A. Swets. Signal Detectability and Psychophysics. New York, NY:Wiley & Sons, 1966.
- [3] D.J. Heeger. Normalization of cell responses in cat striate cortex. *Vis Neurosci* 1992;9:181-97
- [4] L. Itti L, J. Braun, D.K. Lee and C. Koch. A Model of Early Visual Processing. In: *Advances in Neural Information Processing Systems, Vol. 10*, Jordan MI, Kearns MJ and Solla SA (eds.), pp. 173-179, Cambridge, MA:MIT Press, 1998
- [5] S. Nelson, L. Toth, B. Sheth and M. Sur. Orientation selectivity of cortical neurons during intracellular blockade of inhibition. *Science* 1994;265(5173):774-77
- [6] A. Pouget, K. Zhang, S. Deneve and P.E. Latham. Statistically Efficient Estimation using Population Coding. *Neur Comp* 1998;10:373-401

- [7] W.H. Press, S.A. Teukolsky, *et al.* Numerical Recipes in C. Cambridge, MA:Cambridge University Press, 1992
- [8] W.R. Softky and C. Koch. Cortical cells should fire regularly, but do not. *Neural Comp* 1992;4:643-6
- [9] H.R. Wilson, D.K. McFarlane and G.C. Phillips. Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vis Res* 1983;23:873-82.
- [10] H.R. Wilson and R. Humanski. Spatial frequency adaptation and contrast gain control. *Vis Res* 1993;33(8):1133-50
- [11] B. Zenger and D. Sagi. Isolating excitatory and inhibitory nonlinear spatial interactions involved in contrast detection. *Vis Res* 1996;36(16):2497-2513.

Acknowledgments: This research was supported by NSF, NIMH and ONR.



Laurent Itti is a Ph.D. candidate in computation and neural systems at the California Institute of Technology, Pasadena, California. He received his M.S. in electrical engineering in 1994 from the Ecole Nationale Supérieure des Télécommunications, Paris, with a specialization in image processing. His main research interest is in the computational understanding of early visual processing in primates, both from bottom-up and top-down perspectives. His homepage is <http://www.klab.caltech.edu/~itti>.



Christof Koch is professor of computation and neural systems at the California Institute of Technology, Pasadena, California. His research focuses on biophysical mechanisms for storing and processing information in single neurons and small neuronal circuits, as well as the computations underlying motion and visual attention in mammalian cortical networks. Together with Francis Crick, he works on the neuronal basis of visual awareness and consciousness. His homepage is <http://www.klab.caltech.edu>.



Jochen Braun received B.S. degrees in Chemistry and Biochemistry in Tuebingen, Germany, and a Ph.D. in Biophysics from UC Berkeley. (1985). Changing his field of research to primate vision, he worked at the Weizmann Institute of Science, Israel, and the California Institute of Technology. His overall goal is to understand "mid-level vision" at the behavioral, computational, and neural circuit levels. He defines mid-level vision to include visual attention, awareness, grouping, and plasticity, but not visual recognition and memory.