Looking for hallmarks of generative models in the visual cortex

Gergő Orbán[1], Pietro Berkes[2], Máté Lengyel[3], József Fiser[1]

1, Volen Center for Complex Systems, Brandeis University; 2, Gatsby Computational Neuroscience Unit, University College London; 3, CBL, Department of Engineering, University of Cambridge

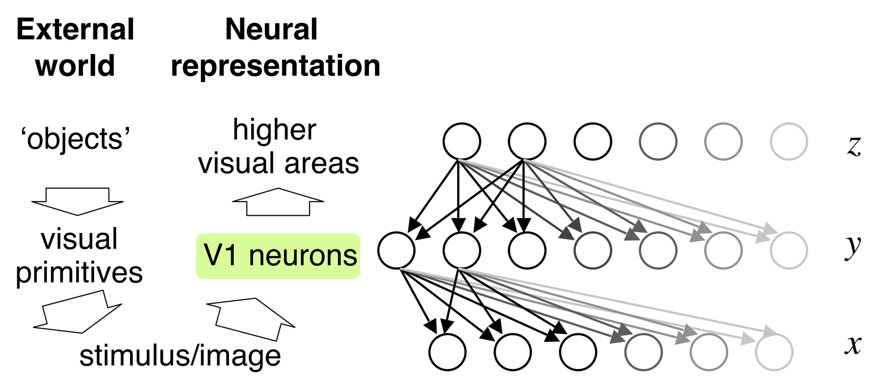
Introduction

The *generative modeling* or *analysis-by-synthesis* approach is a probabilistic framework to understand perception [9,14]. One major challenge for assessing the relevance of this framework to cortical function is to establish if and how the neural hardware implements such probabilistic computations. We propose a novel approach to this problem that, rather than testing a specific generative model, looks for hallmarks in neural activity that are fundamental features of the entire class of generative models. By this approach, we can relate physiological data to model predictions regardless of the particular features of the model under assessment.

Framework

GENERATIVE MODELING HYPOTHESIS

- The visual system embodies a probabilistic model of how visual elements (y, z) combine to form an image (x): $P(x|y)$, $P(y|z)$, $P(z)$



Visual perception

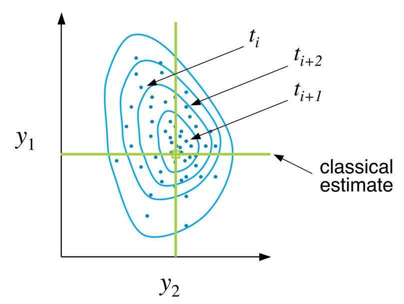
- Inverting the model to infer likely causes of the stimulus: $P(y|x)$
- Result of inference: ideally, not a single estimate but a *posterior probability distribution*

$\Longrightarrow$ A normative, functional account of vision:

- Accounts for electrophysiological and psychophysical data (context-dependency, attention, illusions) [5,7,10]
- Reproduces a variety of RF characteristics of V1 neurons [2,6,8,11,13]

SAMPLING HYPOTHESIS How can a distribution be represented by neural activity?

Sampling hypothesis [5]



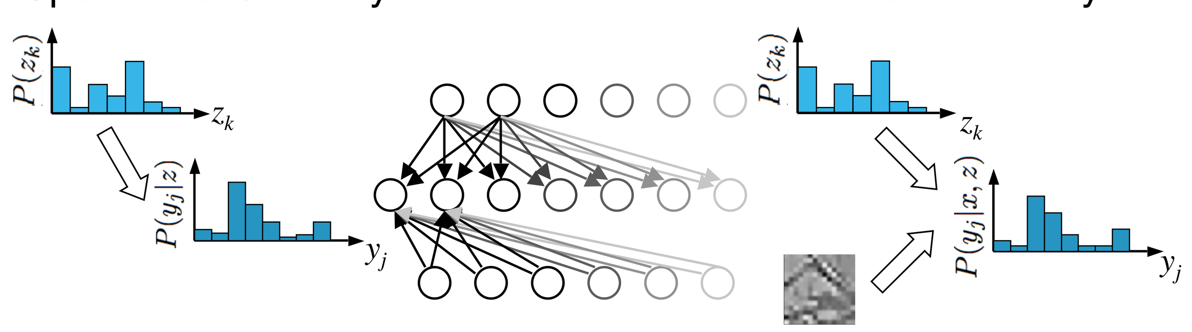
- Compatible with traditional interpretation: classical view relied on the marginals of the distribution (or the mean of it)
- Consistent with the observation of high trial-by-trial variability [1,4,12]

How can generative models be traced down by experiments?

- The space of possible generative models is extremely large
- The 'exact' generative model might be too complex

OUR APPROACH:
Look for hallmarks in neural activity that are characteristics of the entire class of models

CURRENT FOCUS: Spontaneous activity

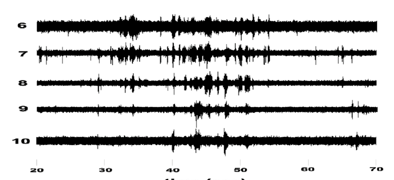


How are these activities related?

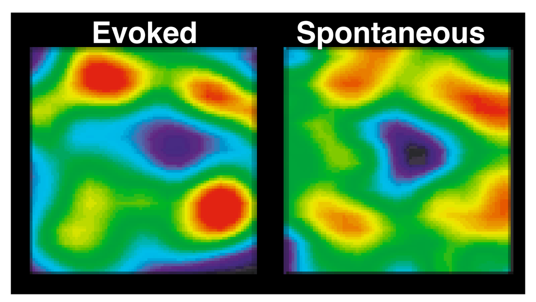
$$\underbrace{P(y)}_{\text{spontaneous activity}} = \int \underbrace{P(y|x)}_{\text{evoked activity}} \; \underbrace{P(x)}_{\text{stimulus statistics}} \, dx$$

Neural data

- In the awake brain there is structured neural activity not directly related to the stimulus

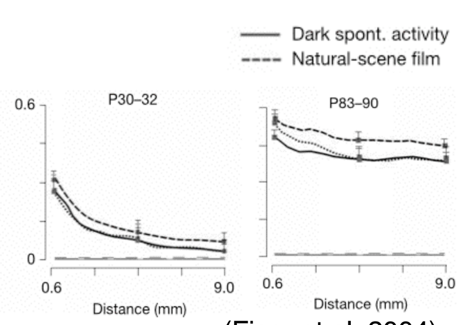


- Structure of neural activities are similar in stimulus evoked condition and closed eye condition



(Tsodyks et al, 1999)

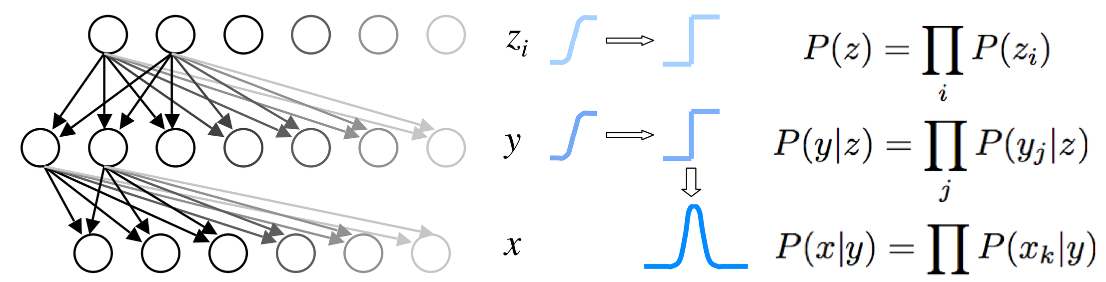
- There are long-range correlations in neural activity



(Fiser et al, 2004)

Model

In silico we can assume a specific generative model and test our hypothesis under experimental conditions



Evoked activity (EA)	$P_{\mathrm{EA}}(y\|x)$	$\int P(y\|x,z)\,P(z)\,dz \propto \int P(x\|y)\,P(y\|z)\,P(z)\,dz$
Marginalized EA	$P_{\mathrm{mEA}}(y)$	$\int P_{\mathrm{EA}}(y\|x)\,P(x)\,dx$
Spontaneous activity (SA)	$P_{\mathrm{SA}}(y)$	$\int P(y\|z)\,P(z)\,dz$

LEARNING THE MODEL

- Pretraining with greedy RBM [15]
- Performing Expectation-Maximization on natural image patches
  1. *Expectation:* Gibbs sampling from the posterior

$$P\left(y_1|y_{\{n\}}, z, x\right) = \frac{P(x|y)P(y_1|z)}{\sum_{\{y_1\}} P(x|y)P(y_1|z)}; \quad P(z_1|z_{\{n\}}, y, x) = \frac{P(y|z)P(z_1)}{\sum_{\{z_1\}} P(y|z)P(z_1)}$$

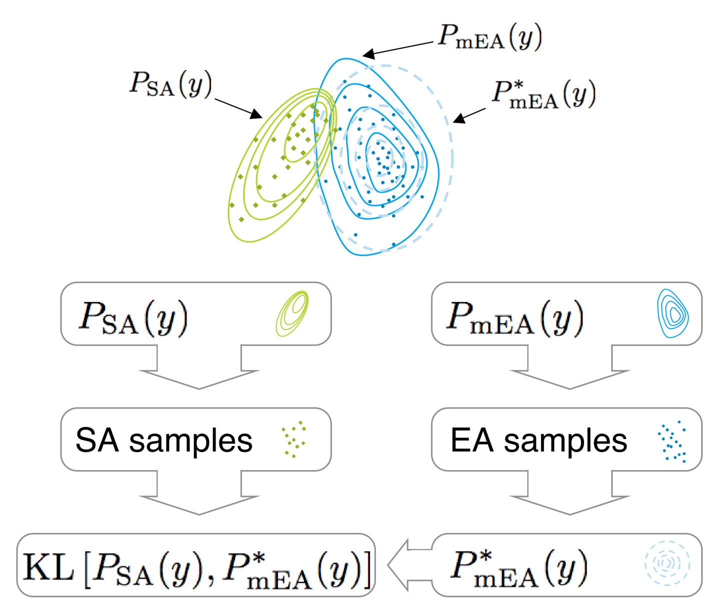
  2. *Maximization:* $\theta^* = \mathrm{argmax}_\theta \prod_{\{x\}} P(x|\theta)$

Evaluation of differences in SA and EA distributions

Küllback-Leibler divergence (KL) is a principled tool for providing a measure of simarity between distributions:

$$\mathrm{KL}\left[P_{\mathrm{SA}}(y), P_{\mathrm{EA}}(y)\right] = -\int P_{\mathrm{SA}}(y) \log P_{\mathrm{EA}}(y)\,dy - H(P_{\mathrm{SA}}(y))$$

Cartoon of KL calculation



- We only have samples from the distributions
- Samples from the SA can be used to perform a Monte-Carlo integral:

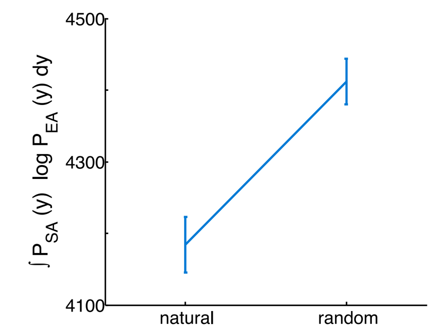
$$\int P_{\mathrm{SA}}(y) \log P_{\mathrm{EA}}(y)\,dx = \frac{1}{N_{samples}} \sum_{y_i \in P_{\mathrm{SA}}} \log P_{\mathrm{EA}}(y_i)$$

- Samples from EA are used to learn a distribution

Results

- We compare EA and SA distributions under different conditions

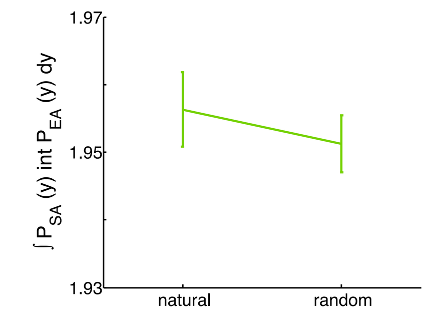
KL for natural and random image-evoked activities



Similarity is significantly higher for EA on natural images

- Difference is expected to be smaller in a naïve model (young animal)

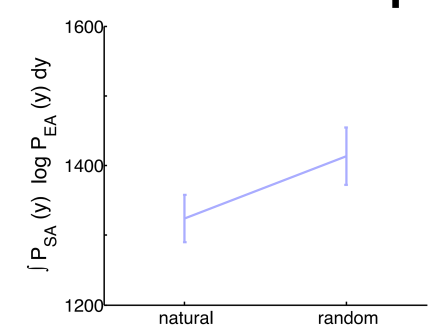
KL in untrained network



Naïve model can not distinguish between natural and random image patches

- Under experimental conditions, only a subset of neurons can be accessed

KL on partial observation



Partial data is sufficient for evaluating the EA and SA distributions

Conclusions

- Variability in evoked activity and spontaneous activity have a common interpretation in the framework of Bayesian generative models
- Distribution of spontaneous activity is related to the distribution of evoked activity through marginalizing over the image statistics
- We demonstrated a tool for evaluating distributions of spike train data: KL divergence can be used for the characterization of the whole distribution of neural activities even if only a few neurons can be measured

References
[1] Arieli, A. et al. (1996) Science, 273, 1868-1871.
[2] Berkes, P., et al. (2007). Cosyne 2007, Salt Lake City, II-111
[3] Blake, R. and Logothetis, N.K. (2002) Nature Reviews Neuroscience, 3:13-21.
[4] Dean, A.F. (1981) Exp Brain Res, 44:437-440.
[5] Hoyer, P.O. and Hyv?rinen, A. (2003) In: Adv in Neural Inf Proc Systems 15, pp. 277-284, MIT Press.
[6] Karklin, Y. and Lewicki, M.S. (2005) Neural Computation, 17(2):397-423.
[7] Lee, T.S., and Mumford, D. (2003) J Opt Soc of America A, 20(7):1434-1448.
[8] Olshausen, B.A. and Field, D.J. (1996) Nature, 381(6583):607-609.
[9] Rao, R., Olshausen, B.A., and Lewicki, M.S., Eds. (2002) Probabilistic Models of the Brain: Perception and Neural Function. MIT Press.
[10] Rao, R. (2005) Neuroreport, 16(16), 1843-1848.
[11] Rehn, M. and Sommer, F.T. (2007) J Comp Neurosci, 22(3), 135-146.
[12] Tolhurst, D.J., et al. (1983) Vision Research, 23:775-785.
[13] Wainwright, M.J. and Simoncelli, E.P. (2000) In Adv Neural Inf Proc Systems 12, pp. 855-861, MIT Press.
[14] Yuille, A.L. and Kersten D. (2006) Trends Cogn Neurosci, 10(7), 301-308.
[15] Hinton et al. (2006) Neural Comput. 18, 1527-1554

Supported by the Gatsby Charitable Foundation IST-FET-1940 program and the Sloan-Swartz Foundation