A generative model for dynamic sensory scenes, and corresponding inference in a neural layer

Timm Lochmann, Udo Ernst, Sophie Deneve

October 11, 2010

[Figure 1 about here.]

The proposed generative model for a dynamic sensory scene is shown in fig.1. The presence or absence of each object in the scene if formalized by a hidden Markov model (HMM) with binary hidden states $X_i^t \in \{0, 1\}$ and binary observations $S_j^t \in \{0, 1\}$. $X_i^t = 0$ and $X_i^t = 1$ are called the 'off-state' and the 'on-state' and for the S_j^t, 0 vs. 1 stands for a spike vs. no spike from presynaptic neuron j, respectively.

During small time intervals Δt, the probabilities of the state X_i^t to switch from 0 to 1 and vice versa are given by $r_i^{\mathrm{on}} \Delta t = P(X_i^{t+\Delta t} = 1 | X_i^t = 0)$ and $r_i^{\mathrm{off}} \Delta t = P(X_i^{t+\Delta t} = 0 | X_i^t = 1)$. r_i^{on} and r_i^{off} control respectively the rate of appearance and the average duration of the stimulus.

the probability of observing a spike in channel j in time interval Δt given the configuration $X^t := [X_1^t, X_2^t, ..., X_M^t]$ of hidden objects is modeled as a linear superposition:

$$p(S_j^t = 1 | X^t) \quad = \quad \Delta t \left(q_0 + \sum_i X_i^t q_{ij} \right) . \tag{1}$$

where $\Delta t \, q_{ij}$ stands for the probability that object i causes a spike in receptor j in an interval of length Δt. The term $\Delta t \, q_{0j}$ models the effect of unspecified causes such as background noise. In the limit of small Δt, the instantaneous firing rate of receptor j is given by $q_{0j} + \sum_i X_i^t q_{ij}$.

Derivation without overlap between objects

Let us first consider the case when there is no overlap between predictive fields, i.e. each object i affects a distinct set of receptors. In this situation, if q_{ij} is larger than 0, then $q_{kj} = 0$ for all $k \neq i$. This situation does not require explaining away, and is equivalent to I independent HHMs, one for each object.

Let $p_i^t = p(X_i^t = 1|S^{(t)})$ denote the probability of feature i being present at time t given synaptic input $S^{(t)} := [S_1^{(t)}, S_2^{(t)}, ..., S_M^{(t)}]$ with $S_j^{(t)} := [S_j^1, S_j^2, ..., S_j^t]$ being the input from receptor units j up to time t. The dynamics of unit i are described via the *log odds* $L_i^t := \log \frac{p_i^t}{1-p_i^t}$. As a consequence, the probability of feature i being present is a simple function of L_i^t, i.e. $p_i^t = [1 + \exp(-L_i^t)]^{-1}$.

Taking the limit of $\Delta t \to 0$ for the discrete time HMM yields a continuous process with temporal dynamics $\dot{L}_i = \frac{d}{dt} L_i$ given as

$$\dot{L}_i = r_i^{\text{on}} \left(1 + e^{-L_i}\right) - r_i^{\text{off}} \left(1 + e^{L_i}\right) + \sum_j w_{ij} s_j - \Psi_i \tag{2}$$

where $s_j = \sum_k \delta(t_{jk} - t)$ refers to the input spike train from channel j. A derivation of this result is provided in Deneve (Denève, 2008).

The weights w_{ij} for incoming spikes and the drift term Ψ_i are given as:

$$w_{ij} := \log \frac{q_{ij} + q_{0j}}{q_{0j}} \qquad \text{and} \qquad \Psi_i := \sum_j q_{ij}. \tag{3}$$

Outputs are generated by integrating eq. (2) up to a dynamic threshold G_i whose dynamics read

$$\dot{G}_i = r_i^{\text{on}} \left(1 + e^{-G_i}\right) - r_i^{\text{off}} \left(1 + e^{G_i}\right) + \eta o_i - \Upsilon_i \tag{4}$$

where $o_i = \sum_k \delta(\tilde{t}_{ik} - t)$ is the output spiketrain of unit i. The unit is said to fire spikes at times $\tilde{t}_{ik}$ when L_i exceeds G_i by more than $\eta/2$ at which point the threshold is increased by η. Υ_i is a constant drift term analogous to Ψ_i. Using the abbreviation $\Phi_i(G) := -r_i^{\text{on}}(1 + e^{-\mathcal{G}}) + r_i^{\text{off}}(1 + e^{\mathcal{G}}) + \Upsilon_i$ one can derive equation (**??**) for G.

L_i^t estimates the probability of object i being present given the receptor inputs. This is an analog quantity, but it has to be signalled via a binary output spike train O_i^t. The adaptive threshold G_i 'simulates' the dynamics of the internal probability estimate of a putative postsynaptic unit receiving O_i^t as input. Whenever G_i decays too far below the actual probability L_i, unit i fires a new spike such that a putative postsynaptic unit can appropriately update its probability estimate. As a result, the dynamic threshold tracks the probability of presence of the object. η controls the precision of this spike-based representation of probability, and regulates the number of output spikes. Alternatively, this neuron can be understood as an integrate and fire neuron with membrane potential V_i^t corresponding to the 'prediction error' $V_i^t = L_i^t - G_i^t$. The threshold and reset potential of this integrate and fire unit are respectively $\frac{\eta}{2}$ and $-\frac{\eta}{2}$. The advantage of using such spike-based rather than a rate-based representation of probabilities are described elsewhere (Denève, 2008).

Derivation of Input targeted Divisive Inhibition (overlap)

We now extend the previous results to a network that can account for different causes i. Analogous to eq. (3), evidence for object i, observed in channel j, (spikes from receptor j) should be weighted by the log ratio of firing rates when the object is present vs. absent.

In contrast to the case of a single hidden cause, these firing rates now depend on the current presence or absence of all other objects $k \neq i$:

$$q_{0j} \longrightarrow q_{0j} + \sum_{k \neq i} X_k^t q_{kj}$$

$$q_{0j} + q_{ij} \longrightarrow q_{0j} + q_{ij} + \sum_{k \neq i} X_k^t q_{kj}$$

Although the true state of the hidden causes is not known to the network, approximate inference can still be implemented via a mean field approach. We use the fact that $\hat{p}_i^t := 1/(1 + e^{-G_i^t})$ is a good estimate of the posterior probability or expected state $p_i^t = 1/(1 + e^{-L_i^t})$.

We thus replaced the binary X_k^t for all but unit i by their on-line estimate $\hat{p}_k^t$. This yields the expected firing rate for channel j when object i is absent vs. present as:

$$q_{0j} + \sum_{k \neq i} \hat{p}_k^t q_{kj} \quad \text{and} \quad q_{0j} + q_{ij} + \sum_{k \neq i} \hat{p}_k^t q_{kj} \tag{5}$$

With the abbreviation $A_{ij}^t := q_{0j} + \sum_{k \neq i} \hat{p}_k^t q_{kj}$ referring to the influence of causes other than X_i^t, the probabilities of observing NO event in absence vs. presence of cause i are

$$1 - \Delta t A_{ij}^t \quad \text{and} \quad 1 - \Delta t (q_{ij} + A_{ij}^t). \tag{6}$$

This provides all the information necessary for the inference algorithm and yields the discrete-time multi-unit equivalent of eq. (2):

$$L_i^{t+\Delta t} \approx L_i^t + \Delta t \left[r_i^{\mathrm{on}} \left(1 + e^{-L_i^t}\right) - r_i^{\mathrm{off}} \left(1 + e^{L_i^t}\right) \right] + \sum_j \tilde{w}_{ij}^t S_j^t + \sum_j b_{ij}^t (1 - S_j^t) \tag{7}$$

with

$$\tilde{w}_{ij}^t := \log \left(\frac{q_{ij} + A_{ij}^t}{A_{ij}^t} \right) \qquad \text{and} \qquad b_{ij}^t := \log \left(\frac{1 - \Delta t (q_{ij} + A_{ij}^t)}{1 - \Delta t A_{ij}^t} \right). \tag{8}$$

In the limit of $\Delta t \to 0$ equation (7) gives the continuous equation

$$\dot{L}_i = \Phi_i(L_i) + \sum_j \tilde{w}_{ij}(t) s_j. \tag{9}$$

It can be shown that this inference algorithm is exact as long as (1) $\hat{p}_i^t = p_i^t$ and (2) the probability that more than 3 objects are present at the same time and overlap on the same receptor can be neglected. Thus, we expect this algorithm to be particularly efficient when objects are rare events, i.e. if $r_i^{\mathrm{on}} \ll r_i^{\mathrm{off}}$.

Equation (8) realizes a type of *divisive inhibition* as can be seen most clearly for $q_{ij} \ll q_{0j}$, i.e. when input weights are small. We can then use fixed weights $w_{ij} := \log \frac{q_0 + q_{ij}}{q_0}$ and a more standard type of divisive inhibition to approximate the effective weights $\tilde{w}_{ij}^t$ in (8) as $\tilde{w}_{ij}^t = \frac{w_{ij}}{1 + \sum_{k \neq i} w_{kj} \hat{p}_k^t}$.

It is important to note that the feedforward weights w_{ij} are fixed and determined by the parameters of the causal model. The effective weights $\tilde{w}_{ij}$, however, depend on the network activity and are therefore not static.

An example of sensory input $S^{(t)}$ generated by the generative model, and outputs of the neural layer is shown on fig**??**.

[Figure 2 about here.]

References

Denève, S. (2008). Bayesian spiking neurons I: Inference. *Neural Computation 20*(1), 91–117.

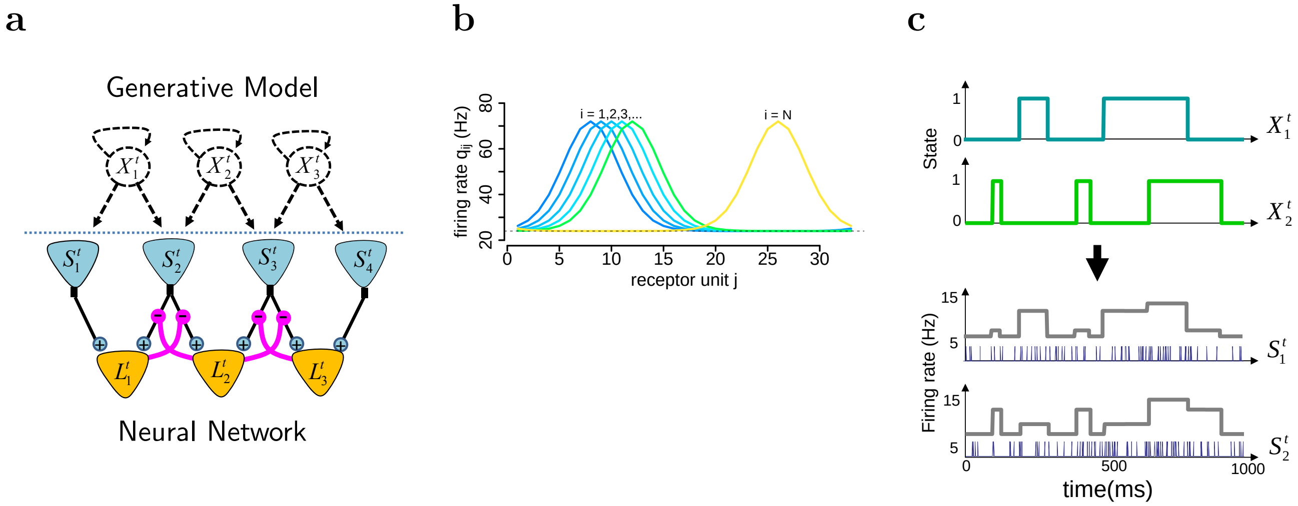

Figure 1: The generative model and its neural implementation. **(a)** Relation between GM and neural network. The dashed circular units represent objects composing the sensory scene, and dashed loopy arrows indicate their stochastic 'on' and 'off' transitions. Straight dashed arrows indicate that these objects modulate in turn the firing rate of receptor neurons (blue triangles) producing spike trains (observations) S_j. The neural network (below the dashed line) is composed of the receptor neurons (input layer) and detector neurons (output layer, represented by orange triangles). Detector neurons process their inputs using feed-forward (black) and inhibitory lateral (magenta) connections. Lateral connections modulate the gains of feed-forward connections (magenta circles) and thus regulate the flow of information between the two layers. **(b)** Predictive fields used as a simplified model of object structure. Each colored line represents the profile of increased firing rate of receptor units caused by one object (color coded). **(c)** An example illustrating how different objects cause correlated and noisy receptor responses. Two objects X_1 and X_2 appearing and disappearing over time (top 2 panels) both influence the time dependent firing rates (i.e. the probability of firing) of two receptor units S_1 and S_2 (bottom 2 panels, plain lines). While the presence of X_1 has a stronger impact on S_1 (i.e. $q_{11} > q_{12}$), X_2 has a stronger impact on S_2 (i.e. $q_{22} > q_{21}$). Spikes from S_1 and S_2 (blue vertical lines) are samples from these rates.

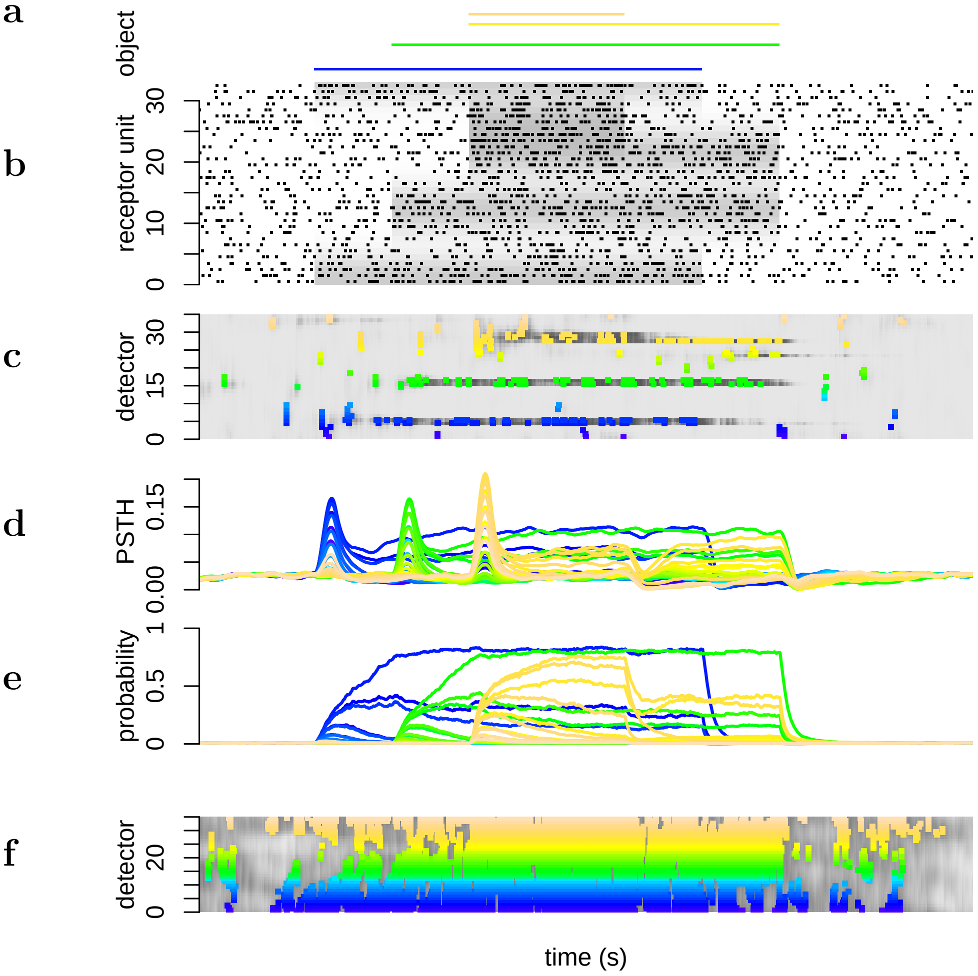

Figure 2: Processing with and without divisive inhibition. **(a)** Presence of objects. Each line corresponds to an object and its presence is indicated by a colored rectangle. The firing rates of the receptor units are determined by the configuration of objects at each point in time. **(b)** Raster plot of input spike trains. Firing rates of corresponding receptor units are indicated by shades of gray. **(c)** Raster plot of output spike trains from the network. Gray shading in the background indicates estimated probabilities of the corresponding units. **(d)** Post stimulus time histogram indicating estimated firing rates over 500 repetitions. **(e)** Probability of object being present decoded from the output spike trains, averaged over 500 repetitions. **(f)** Raster plot for the network without divisive inhibition.