

Spike-latency coding of topological feature homogeneity, and the shaping of synaptic potentials by forward inhibition

Rüdiger Kupper, Marc-Oliver Gewaltig, Andreas Knoblauch, Ursula Körner, Edgar Körner

2005

Preprint:

This is an accepted article published in Proceedings of the 3rd HRI International Workshop on Advances in Computational Intelligence. The final authenticated version is available online at: https://doi.org/[DOI not available]

SPIKE-LATENCY CODING OF TOPOLOGICAL FEATURE HOMOGENEITY, AND THE SHAPING OF SYNAPTIC POTENTIALS BY FORWARD INHIBITION

Rüdiger Kupper, Marc-Oliver Gewaltig, Andreas Knoblauch, Ursula Körner, and Edgar Körner

Honda Research Institute Europe GmbH Carl-Legien-Str. 30, D-63073 Offenbach/Main, Germany

ABSTRACT

Spiking neurons can code stimulation strength in firing latency. This latency code carries the same information as firing rate, but is readily established in a few milliseconds. The brain can employ this fast and accurate time-based code for the processing of homogeneous stimulation. We have proposed that the koniocellular path of the visual system performs spatial homogeneity detection. Here, we present a model implementation of this process, using spike-latency code. In a network simulation, we demonstrate the fast detection of spatially homogeneous luminance. We also demonstrate, how the homogeneity threshold can be dynamically adjusted. Here, we use forward inhibition to shape synaptic potentials.

1. INTRODUCTION

Spiking neurons can code stimulation strength in firing latency [1]. Figure 1 shows a recording from orientationselective cells in the primary visual cortex, taken from the original works by Hubel and Wiesel. Latency of the first response spike is large when firing rate is small, and vice versa. Both carry the same information, but the latency code can readily be assessed after a few milliseconds, while readout of rates affords integrating over time or space.

In a recent publication [2], we described neural mechanisms for putting ensembles of spiking neurons into consistent internal states. Such ensembles then use a common time frame for latency coding. The brain can then employ this fast and accurate time-based code to detect homogeneous stimulation in a group of neurons. We have previously suggested homogeneity detection ("surface detection") to support stimulus processing in primate visual cortex [3]. Here, we demonstrate in a network simulation, how visual neurons can detect spatially homogeneous luminance. The latency code can be flexibly read-out for this purpose. We show how a simple neural process (forward inhibition) can dynamically adjust the threshold, at which homogeneity is detected.



Fig. 1. Firing rate and response latency of a typical V1 neuron, in response to an oriented bar stimulus. Left: Firing rate as a function of bar orientation. Right: Latency of first spike in response to four different bar orientations (marked red). Response latency and firing rate carry the same information: Latency of the first response spike is large when firing rate is small, and vice versa.

2. TOPOLOGICAL FEATURE HOMOGENEITY

Due to their internal dynamics, spiking neurons act as *co-incidence detectors* [4, 5]. A packet of input spikes converging onto a neuron increase the probability of a response spike, if they arrive in a narrow time window, in contrast to being dispersed over a long period of time. At the same time, in the framework of latency coding, coinciding action potentials have a distinct meaning: If neural elements reliably convert stimulus strength into firing latency, then coincidentally firing neurons must have received the same stimulation. If these neurons prefer a certain stimulus feature, this feature must have been similarly present in their receptive fields (fig. 2).

This principle can be applied to topologically arranged sets of neurons. This creates, from a set of detectors for feature f, a set of detectors for homogeneous appearance of f (fig. 3). The principle requires only two well-known architectural ingredients, and they can be found virtually everywhere in cortex: Topologically arranged sets of feature detectors, and topology-preserving convergent projections.



high variance – asynchronous spikes

low variance - synchronous spikes

Fig. 2. The principle of spike-latency based homogeneity processing. Example: A set of visual neurons respond at latencies depending on the luminance inside their receptive field. The receiving neuron responds when their action potentials coincide. This makes it a detector for homogeneous luminance.

2.1. The generic design pattern for feature homogeneity detection

The above described design principle can be formalized in the following way:

- 1. A sending neural population N_s is a set of latencycoding feature detectors for feature f. (That is, firing of a neuron in N_s corresponds to the appearance of feature f, and latency corresponds to some gradual quality q of f, usually its strength. However, q may correspond to any feature quality that can be mapped to latency, such as size or orientation.)
- The neural population N_s is arranged across the cortical surface preserving the topology of a stimulation space T (e.g. retinotopically, tonotopically, somatotopically, etc.).
- 3. Neurons in a receiving population N_r receive convergent input from local sub-populations of N_s with a fixed diameter d measured in **T**, and with a fixed transmission delay.
- The local sub-populations of N_s and respective receiving neurons in N_r are chosen as to preserve topology. (This makes N_r topologically arranged according to T.)

In this case, activity of a neuron in population N_r indicates the appearance of feature f with homogeneous quality q across a local region of diameter d in the stimulation space \mathbf{T} , and at the location corresponding to the respective neuron's topological position. Moreover, the quality q of appearance of feature f is encoded in the firing latency of this neuron.

This is a generic design principle, which in general is completely independent from the kind of stimuli or features that shall be processed. Visual, auditory and somatosensory



Fig. 3. Detecting feature homogeneity from spike latency in a topological arrangement of feature detectors. Homogeneity detection is implemented in two network stages. N_s is a topologically arranged set of feature detectors for feature f. They generate a spike-latency code for the quality of feature f at the respective location in the input space. **Left:** A single N_r neuron receives action potentials from a local patch of N_s neurons. Its sensitivity for coincident synaptic events makes this neuron a detector for homogeneous appearance of feature f inside its receptive field. **Right:** Each neuron of the complete N_r population receives action potentials from a topologically corresponding local patch of N_s neurons. This makes N_r a topologically arranged set of detectors for homogeneous appearance of feature f.

features come to mind, suggesting that this design principle may be re-used at various cortical locations.

2.2. Methods

In the remainder of this paper, we will use a concrete implementation of the above described generic design pattern, giving a *proof of principle*, while at the same time allowing for further investigation. We choose an example from the visual domain.

In [3], we suggested that the konio-cellular pathway of the visual system plays an important role in the processing of surface-like regions in the visual stimulus. "Surfacedetecting" konio-cells could improve the reliability of edgedetection. They can neurally impose the principle, that "where's a surface, there can't be edges". "Surface-detection" is supposedly done by konio-cells with large receptive fields, and lacking an antagonistic surround. Cells with these properties have been observed in macaque cortex [6].

Network

Our model consists of two retinotopically arranged sets of spiking neurons, applying the generic design pattern described above (fig. 3). The lower network stage (N_s) is a retinotopic set of latency-coding feature detectors, which are sensitive to local luminance. The higher network stage (N_r) applies the principle of spike-latency based homogeneity detection to the output of these neurons. It corresponds to a retinotopic set of konio-cells with large receptive fields and no antagonistic surround.

For stimulating our simulated network, we select patches of 100×100 pixels from natural gray-scale images (fig. 5, panel A). The local gray values of these stimulus patches are transformed into neural spike responses by injecting currents of corresponding magnitude into a 100×100 layer of model integrate-and-fire neurons (N_s) . The mapping of image pixels to stimulated neurons is retinotopic. In order to generate a latency code from this stimulation, the N_s neurons need to be suitably prepared. To initiate latencycoding, we use the methods analyzed in [2] (artificial reset, input suppression, common inhibition). They put the N_s neurons into consistent internal states. This ensures reliable conversion of stimulation strength into spike latency, using a common time-frame for latency coding across the whole population. Permanently applying the stimulus, we initiate latency coding, and allow the neurons to fire for 100 ms. We then repeatedly re-initiate latency coding after each 100 ms of firing. This makes population N_s a retinotopically arranged set of latency-coding neural detectors for local luminance (the feature f). The brighter the visual stimulus in their receptive field (feature quality q), the shorter the latency.

As a next processing stage, sub-populations of diameter d = 11 neurons of N_s (representing $d = 1.1^{\circ}$ visual angle) project convergently onto another layer of 100×100 model neurons (N_r) , while preserving topology (fig. 3). Transmission delays are identical for all connections. According to the design principle described above, this makes population N_r a retinotopically arranged set of neural detectors for *spatially homogeneous* luminance across distances of $d = 1.1^{\circ}$ visual angle.

We use the NEST simulator developed in collaboration with the Neural Simulation Technology Initiative [7] for simulating the structured neural network. During the whole simulation, we record the spike responses of all neurons, and then analyze this data with respect to response latencies and robustness of spike production.

2.3. Results

We recorded the action potentials of N_s and N_r neurons, mainly being interested in their *response latency*. We define response latency as the time until a neuron's first action potential, after latency-coding has been initiated.

Figure 4 shows spike-trains recorded from N_s and N_r neurons. These are typical responses. N_s neurons have been excited according to the visual stimulus patch shown in panel A, and latency coding has been initiated at t = 0, using three different methods (reset, input suppression, common inhibition [2]). Regions of homogeneous luminance in the stimulus patch cause groups of neighboring N_s neurons to respond with similar latencies. Their action potentials show up as "spike-fronts" in the plots. These coincident spikes, in turn, cause responses in the N_r neurons in corresponding locations.

Figure 5 indicates typical response latencies of N_r neurons on a gray-level scale. Blue color indicates, that the respective neuron did not produce an action potential. Panel A shows three different stimulus patches. Panel B depicts response latencies obtained, when the N_s neurons are optimally prepared for latency coding. This is achieved by an artificial reset of the model neurons, a process that we use for evaluation, but which cannot occur in the real brain. Panels C and D depict response latencies obtained, when the N_s neurons are prepared by two biologically plausible mechanisms, input suppression and inhibition [2]. They can achieve only a partial reset of the neurons' internal states. This affects the fidelity of latency coding, and consequently, response latencies jitter.

Still, locations of active N_r neurons correspond to regions of homogeneous luminance in the stimulus patch the N_r neurons act as *homogeneity detectors*. By relying on only the first action potentials of latency-coding neurons, spike-latency based homogeneity processing is very fast, with first components signaled after 5–20 ms.

3. CONTROLLING THE HOMOGENEITY CONSTRAINT BY FORWARD INHIBITION

Spike-latency based homogeneity detection relies on the constructive superposition of post-synaptic potentials (PSPs). The PSP caused by an incoming action potential is usually modeled as an alpha function (fig. 6, red curve),

$$U(t) = t \ e^{-\frac{t}{\tau}},$$

the sign and time constant depending on the type of synapse that is to be modeled (inhibitory, excitatory, AMPA, NMDA, etc.). The function has a hill-shaped appearance, reaching its maximum at $t = \tau$ and then declining.

For a set of excitatory spike events to raise the receiving neuron's membrane potential above firing threshold, the events must be close enough in time for their PSPs to overlap and thus superimpose constructively. The allowed temporal spread of synaptic events is determined by the synaptic time constant τ . The smaller τ , the smaller the overlap of PSPs, and the greater the required coincidence of incoming



Fig. 4. Processing of spatial homogeneity. Typical spike responses of $100 N_s$ and N_r neurons. Topological positions of the 100 selected neurons correspond to the center row of pixels in the stimulus patch (marked in red). Spike time histograms shown in gray at the bottom of spike trains. A: Stimulus patch. B, E: Spike responses of N_s and N_r neurons, when N_s neurons were prepared for latency coding using artificial reset. C, F: Using input suppression. D, G: Using inhibition.



Fig. 5. Detecting homogeneity of luminance. Results from our network simulation. A: Stimulus patches. **B–D:** Typical response latencies in the neural population N_r . (Blue: no action potential was produced.) **B:** N_r latencies obtained, when the sending neural population N_s was optimally prepared for latency coding by an artificial reset. **C and D:** N_r latencies obtained, when N_s was prepared by neurophysiologically plausible reset mechanisms (input suppression and inhibition [2]). **E:** Overlay of the stimulus patches and the region classified as homogeneous (taken from panel B). This simulation featured an additional OFF-pathway for the processing of dark regions. Both, ON- as well as OFF-responses are shown in the plots.



Fig. 6. Shaping of an excitatory PSP by delayed feed-forward inhibition. Red: Excitatory PSP with $\tau=2$ ms. Green: Inhibitory PSP with $\tau=2$ ms and delay $\Delta t=1$ ms. Dashed black: Effective PSP resulting from superposition.

action potentials. For spike-latency based homogeneity processing, this translates into a higher amount of homogeneity required in the input features.

Shaping of post-synaptic potentials

It is obvious that controlling the synaptic time constant τ would be beneficial, because this would enable us to select the grade of homogeneity to be detected. However, controlling τ through biologically plausible network effects is difficult. In most cases, the speed of synaptic dynamics is fixed, or depends on chemical processes which cannot be directly influenced for the means of neural computation.

There is, however, the possibility of shaping the *effective PSP* caused by a synaptic event. This is achieved by pairing an excitatory action potential with an inhibitory action potential transmitted shortly afterwards to the same neuron (forward inhibition). Figure 6 shows the effect of an inhibitory PSP arriving shortly after an excitatory PSP of same strength and time constant. The resulting *effective PSP* is much narrower than the original curve, raising the constraint for constructive superposition of subsequent events.

3.1. Methods

Controlling the delay of the paired feed-forward inhibition, the constraints put on feature homogeneity can be gradually changed. This effectively determines the threshold, at which a local region is classified as homogeneous by the N_r neurons. We have included the principle of PSP shaping into our example network for the processing of spatial homogeneous luminance. We extended the architecture by duplicating each connection from the sending to the receiving neuron layer. Action Potentials transmitted along these duplicated connections cause inhibitory PSPs in the postsynaptic N_r neurons. They have a fixed time delay relative to the excitatory action potential (see illustration in fig. 7).



large Δt – homogeneity threshold low

small Δt – homogeneity threshold high

Fig. 7. Controlling the homogeneity threshold, using additional feed-forward inhibition. Each excitatory action potential transmitted to the post-synaptic neuron is followed by an inhibitory action potential to the same neuron, with a fixed relative delay of Δt . Shorter delays require higher homogeneity for the detector neuron to respond. Left: The homogeneity threshold is low, the detector responds to the stimulus. Right: The homogeneity threshold is high, the detector does not respond to the (identical) stimulus.

We then systematically changed the time delay between excitatory and inhibitory action potential, in otherwise identical simulation runs.

3.2. Results

Figure 8 shows response latencies of the homogeneity processing N_r neurons under these conditions. Inhibitory delays used in the simulation increase from upper left to lower right in reading order. All non-blue portions indicate active N_r neurons. This means that luminance values in the corresponding region of the stimulus patch have been classified as homogeneous. The area identified as homogeneous (nonblue) changes monotonically with the inhibition delay. For short delays, fewer regions of the stimulus patch are classified as homogeneous, due to the much sharper effective PSPs caused by these delays.

Note that the delay, at which action potentials are produced in response to a given stimulus, can be influenced through a variety of network effects. Background stimulation may lower the effective firing threshold of neurons, or the neurophysiological effects of attention may promote neural processing. Hence, the criterion for homogeneity detection, the "homogeneity threshold", may be dynamically changed according to the momentary requirements or interests.

4. DISCUSSION

Processing of feature-homogeneity by neural coincidence detection is a process that strongly depends on the dynamics of the synapses. Incoming action potentials cause superimposing excitatory post-synaptic potentials, that must exceed



Fig. 8. Response latencies of homogeneity processing N_r neurons for $\tau=2$ ms and different values of the feed-forward-inhibition delay. Blue: no action potential, gold: latency > 20 ms.

the firing threshold to cause a post-synaptic action potential. The principle of PSP shaping, described in sec. 3, makes use of this property. In principle, the response of a homogeneity detecting neuron conveys different aspects of information.

4.1. Information content of homogeneity detection

Number of incoming action potentials

One statement that can be trivially formulated is that there exists a lower bound on the number of coincident action potentials to raise the receiving neuron's membrane potential above firing threshold. In general, the number of action potentials participating in the code will influence at what threshold and to what resolution feature homogeneity can be detected. A higher number of participating action potentials increases the number of combinatorial possibilities, and allows for more sharply formed PSPs to be used.

Spike response and latency

The spike response of a homogeneity-detecting N_r neuron will carry two components of information:

1. The fact that an action potential was produced *at all* indicates that the received action potentials arrived close enough in time for their PSPs to superimpose constructively. This indicates that the feature homogeneity in the respective stimulus region exceeds a given threshold.

This is a Boolean (binary) statement (yes/no) on topological feature homogeneity.

2. The response *latency* of the homogeneity detecting N_r neuron will reflect the latency of the coincident input spikes. The N_r response will have an additional, constant delay, given by the spike transmission time from N_s to N_r , and the neuron's internal dynamics.

This is a gradual (analog) value, describing feature quality q.

This means that the N_r neurons do not only detect regions of homogeneous feature activation in the stimulus. They also loop through the latency code of the participating neurons, giving information on *how strongly activated* this feature is. Another way is to look upon the N_r neurons as a kind of neural filters. They block spike responses from non-homogeneous regions, and let the latency code from homogeneous regions pass.

4.2. Robustness of homogeneity processing

Latency-based processing of feature homogeneity relies on the generation of good latency codes in the pre-synaptic network stage N_s . However, the quality of the latency code can be subject to changes [8].

As described above, the output of homogeneity processing neurons carries a binary and an analog component of information: (1) Is an action potential produced? (2) If so, at what latency? Both components can be investigated regarding their robustness to changes in the latency code.

Figure 9 shows the robustness of the binary component. The values depicted are probabilities of spike production (regardless of latency) for repeated presentations of the same stimulus patch (panel A). Latency coding each time was initiated at stimulus onset, using three different methods (reset, input suppression, common inhibition [2]). If homogeneity detection was robust against variations in the latency code, spike probabilities should turn into a binary relation (values 0.0 and 1.0 only). An action potential should either always or never be produced for presentation of a certain stimulus patch. Panel B shows that this is the case, when the N_s neurons are optimally prepared for latency coding. This is achieved by an artificial reset of the model neurons, a process that we use for evaluation, but which cannot occur in the real brain. From panels C and D it can be seen that spike responses of the N_r neurons are sensitive to the fidelity of the pre-synaptic latency code. Again, the two biologically plausible methods, input suppression and inhibition, were used for preparing the N_s neurons. They can achieve only a partial reset of the neurons' internal states, making N_s



Fig. 9. Robustness of spatial homogeneity detection, expressed as spike probability of homogeneity-detecting neurons to repeated presentation of the same stimulus patch. A: Stimulus patches. **B–D:** Spike probabilities in the neural population N_r . **B:** N_r spike probabilities obtained, when the sending neural population N_s was optimally prepared for latency coding by an artificial reset: Spike responses are perfectly reproducible. **C and D:** N_r spike probabilities obtained, when N_s was prepared by neurophysiologically plausible reset mechanisms (input suppression and inhibition [2]). This simulation also featured the additional OFF-pathway for the processing of dark regions. Both, ON- as well as OFF-responses are shown in the plots.

response latencies jitter. This, in return, makes N_r spike production less reliable. However, from panels C and D it can be seen that N_r responses are highly reproducible for strongly homogeneous image regions, while reproducibility is degraded for regions of less expressed homogeneity. Hence, the spike probability of homogeneity-detecting neurons degrades gracefully with decreasing homogeneity.

4.3. Influence of delay and inhibition strength in PSP shaping

As can be seen from fig. 6, using paired feed-forward inhibition for PSP shaping does not only influence the width of the resulting PSP, but also its peak value. This, of course, does have an additional effect on the number and coincidence of incoming action potentials required to exceed the firing threshold. The form of the resulting PSP depends on delay, amplitude, and time constants of the paired PSPs. The question, how this translates into constraints on homogeneity processing, shall not be further addressed here.

4.4. Further applications of topological feature homogeneity processing

The detection of spatially homogeneous visual luminance is not the only example for the processing of topological feature homogeneity. Among those that come to mind are the following:

4.4.1. Processing of texture homogeneity

By processing the topological homogeneity of more complex visual features, we can also detect homogeneous appearances of oriented lines, patterns, etc., depending on the feature f that is extracted from the visual stimulus. A possible application is the detection of object surfaces that are not homogeneous in luminance, but of homogeneous texture, e.g. a table surface, a carpet, ground covered by sand or pebbles, treetops, and water surfaces.

4.4.2. Processing of motion homogeneity or change homogeneity (newness)

If f is a feature that is spatio-temporally defined, spatially homogeneous motion can be detected. A first approach would be the extraction of changes in visual stimulus by computation of difference images. Homogeneity processing of this feature would then indicate coherent change (newness) of extended regions in the stimulus, e.g. (dis)appearing objects. A more sophisticated implementation could use directionally sensitive Reichardt detectors for feature detectors. This would allow for detection of coherently moving extended regions in the visual stimulus, such as moving objects or the processing of ego-motion.

5. CONCLUSION

Recently we have shown, that stimulation of neurons in the visual system can reliably be transformed into a fast and accurate latency code [2]. Here we have formulated a generic design principle, which employs this efficient code for the processing of topological feature homogeneity. The design principle is independent from the kind of stimuli or features to be processed. Visual, auditory and somatosensory features come to mind, suggesting that the method may be reused for homogeneity processing at various cortical locations. Applying this design principle, we presented a spiking implementation of the previously proposed mechanism for surface detection in the konio-path of the visual system [3]. We showed that influencing the temporal overlap of successive PSPs by forward inhibition presents a possible mechanism for setting the sharpness of the homogeneity constraint. In spatial homogeneity detection, this translates to a coarser or finer processing of homogeneous regions in the visual stimulus. This can be used by the visual system to switch between whole-scene processing and the processing of visual details at high resolution.

In addition, the principle of PSP shaping opens up an attractive new opportunity for dynamically controlling the mode of operation in a spike-coding network: Two types of stimulus-related information can be conveyed in the train of action potentials produced by a neuron: (1) information conveyed in the rate of firing, and (2) information conveyed in spike-latency relative to a given time-frame. The subsequent processing of this information by post-synaptic neurons relies on two different mechanisms. For the extraction of (2), the exact timing and separability of single synaptic events is crucial, while the extraction of (1) requires integrating large numbers of synaptic events over space or time. PSP shaping, as a means of controlling the temporal integration of synaptic events by the post-synaptic neuron, holds the opportunity to select, through a neural process, which of the two components of information is extracted. A neuron reliably detecting the coincidence of single spike events in the one case (narrow PSPs) will turn into a neuron determining the mean rate of incoming events, regardless of their temporal precision (broad PSPs). Likewise, an ensemble of neurons acting as the source for fast spike-latency based processing of homogeneity in the one case, will turn into a source for the assessment of instantaneous ensemble rate. Active transition from latency to rate processing in the very same neurons presents a possible approach to the unification of both, spike-time and rate codes, in a single network.

6. REFERENCES

- Rufin van Rullen, Rudy Guyonneau, and Simon J. Thorpe, "Spike times make sense," *Trends Neurosci.*, vol. 28, no. 1, pp. 1–4, Jan. 2005.
- [2] Rüdiger Kupper, Marc-Oliver Gewaltig, Ursula Körner, and Edgar Körner, "Spike-latency codes and the effect of saccades," *Neurocomp.*, vol. 65–66C, pp. 189– 194, 2005, Special issue: Computational Neuroscience: Trends in Research 2005 – Edited by E. de Schutter.
- [3] Marc-Oliver Gewaltig, Ursula Körner, and Edgar Körner, "A model of surface detection and orientation tuning in primate visual cortex," *Neurocomp.*, vol. 52– 54, pp. 519–524, 2003.
- [4] Moshe Abeles, "Role of cortical neuron: integrator or coincidence detector?," *Israel Journal of Medical Sciences*, vol. 18, pp. 83–92, 1982.
- [5] Markus Diesmann, Marc-Oliver Gewaltig, and Ad Aertsen, "Stable propagation of synchronous spiking in cortical neural networks," *Nature*, vol. 402, no. 6761, pp. 529–533, 1999.
- [6] Michael P. Sceniak, Dario L. Ringach, and Michael J. Hawken ADN Robert Shapley, "Contrast's effect on spatial summation by macaque V1 neurons," *Nature Neurosci.*, vol. 2, no. 8, Aug. 1999.
- [7] Markus Diesmann and Marc-Oliver Gewaltig, "NEST: An environment for neural systems simulations," in Forschung und wissenschaftliches Rechnen. Beiträge zum Heinz-Billing-Preis 2001, Theo Plesser and Volker Macho, Eds., vol. 58 of GWDG-Bericht, pp. 43–70. Ges. f. wissenschaftliche Datenverarbeitung mbh Göttingen, 2003, http://www.nest-initiative.org.
- [8] Rüdiger Kupper, Marc-Oliver Gewaltig, Ursula Körner, and Edgar Körner, "Spike-latency codes and the effect of saccades," Tech. Rep. HRI-EU Report 04-08, Honda Research Institute Europe GmbH, Carl-Legien-Str. 30, D–63073 Offenbach/Main, Germany, 2004.