Available online at www.sciencedirect.com
SCIENCE DIRECT®

Neurocomputing 57 (2004) 125 – 134

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Sparse spike coding in an asynchronous feed-forward multi-layer neural network using matching pursuit

Laurent Perrinet[a,*], Manuel Samuelides[a], Simon Thorpe[b]

[a]ONERA/DTIM, 2, av. É. Belin, Toulouse 31055, France
[b]Cerveau & Cognition (UMR 5549), 133, rte. de Narbonne, Toulouse 31062, France

Abstract

In order to account for the rapidity of visual processing, we explore visual coding strategies using a one-pass feed-forward spiking neural network. We based our model on the work of Van Rullen and Thorpe Neural Comput. 13 (6) (2001) 1255, which constructs a retinal representation using an orthogonal wavelet transform. This strategy provides a spike code, thanks to a rank order coding scheme which offers an alternative to the classical spike frequency coding scheme. We extended this model to efficient representations in arbitrary linear generative models by implementing lateral interactions on top of this feed-forward model. This method uses a matching pursuit scheme—recursively detecting in the image the best match with the elements of a dictionary and then subtracting it—and which may similarly define a visual spike code. In particular, this transform could be used with large and arbitrary dictionaries, so that we may define an over-complete representation which may define an efficient *sparse spike coding* scheme in arbitrary multi-layered architectures. We show here extensions of this method of computing with spike events, introducing an adaptive scheme leading to the emergence of V1-like receptive fields and then a model of bottom-up saliency pursuit.

Keywords: Vision; Parallel asynchronous processing; Ultra-rapid categorization; Wavelet Hansform; Natural images statistics; Sparse coding

* Corresponding author. Present address: INCM/CNRS (UMR6193), 31, ch. Joseph Aiguier, 13402 Marseille, France
E-mail address: laurent.perrinet@laposte.net (L. Perrinet).

0925-2312/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2004.01.010

1. Toward an efficient dynamical representation

1.1. How to break the code of vision?

Between neuroscience and neuromorphic engineering, our goal is to understand possible spike coding strategies in the central nervous system. In particular, faced with the light influx from the physical world, what are the strategies in a visual system to extract and transmit the relevant features from the eye to the desired output? The physiology of the neurons, the architecture of the visual system and the statistics of the light inputs are as many constraints on the visual system, and a key challenge is to “break” the code of vision.

In particular, the efficiency of a strategy, e.g. for an animal to categorize prey and predators as quickly as possible, is a main evolutionary constraint on the visual system. Experiments of ultra-rapid categorization [7] in humans and monkeys have showed that the visual system could distinguish high-level categories in as short as 150 ms. It suggests to us to move from the analogy of the visual system with classical image processing strategies to more efficient dynamical neural network models.

This paper will at first present an alternative strategy to the classical paradigm which states that analog retinal activity is coded by the spikes’ firing frequency. We will show how to code the analog values solely by the relative rank of their latency, so that the image is coded by a parallel wave of single spikes. But this model is highly constrained by its architecture and we will then present an extension of this method to arbitrary architectures by the implementation of lateral interactions. Finally, this strategy is extended to a model of sparse spike coding in a multi-layer neural network. We show applications to model the visual system and especially the transform in the columnar organisation of the primary visual cortex (V1).

1.2. Analog to spike coding in the retina

As described in Van Rullen and Thorpe [9], let us first define our model retina as a multi-layered structure characterized by a set of neurons, the ganglion cells (GCs), sensitive at different spatial scales to the local contrast of the image intensity detected at the photo-receptors. The neurons are defined by their position and scale as dilated, translated and sampled *Mexican hat* (or difference of Gaussian DOG) filters (see [3, p. 77]). They are placed uniformly over dyadic scales grids, i.e. growing over both axis as powers of 2. The dendrite of a neuron i may be characterized by its weight vector ϕ_i over its receptive field and the activity at the soma of the neuron is the usual dot product:

$$C_i := \langle I, \phi_i \rangle = \sum_{\vec{l} \in \mathcal{R}_i} I(\vec{l}) \,.\, \phi_i(\vec{l}), \tag{1}$$

where $I(\vec{l})$ is the luminosity at pixel $\vec{l}$ and $\mathcal{R}_i$ is here the receptive field of the neuron i. At first, this architecture is tuned to form an *orthogonal wavelet transform* [3] of the image, so that responses of different filters are uncorrelated (i.e. $\langle \phi_i, \phi_j \rangle = 0$ for $i \neq j$).

Thus, a static image is transformed into a multi-scale representation of analog contrasts. Under the assumption of orthogonality, these analog values are theoretically sufficient to reconstruct the image with a *compact* number of active coefficients. Due to its performance, this representation is often used in image processing algorithms.

1.3. A visual spike code in the optic nerve

When presenting an image at an initial time, each neuron of the model integrates the rectified analog contrast information at its soma until it reaches a threshold:[1] it then emits a spike—the more it is activated, the more the spike is fired rapidly—that propagates along the axon and its activity is resetted. Classically, this generates a pattern of spikes whose instantaneous frequency may constitute the image's code.

But the spike code may also be carried by the exact spiking time (or *latency*) of the first spike. To focus on this transient aspect of the neural code, we will consider solely this spike and show how this part of the signal may efficiently carry the neuronal signal. In this extreme case, the code solely consists of the latency for each of the different fibers i which is inversely proportional to the neuron's excitation current, i.e. to the corrected activity. This algorithm defines a coding scheme from an analog matrix to a spike 'wave front' which will travel along the optic nerve. But, how to decode the analog information attached to every spike?

Though biologically highly unrealistic, we will rate the quality of the image reconstruction by the wave front as an upper bound of information transmission. We will therefore record the spike wave generated by this model retina and in our framework, since the wavelet transform is orthogonal, the image may be simply reconstructed by the coefficients' values:

$$I_{\mathrm{rec}} \sim \sum_i |C_i| \,.\, p_i \,.\, \phi_i .$$

This framework defines thus an algorithm for the *progressive reconstruction* of the image by iteratively transmitting coefficients with higher energy first. It achieves perfect reconstruction toward the original static image if the architecture is orthogonal[2] and that the coefficients are known. But how to transmit these analog coefficients' absolute values along the optic nerve using the spike wave?

In fact, Van Rullen and Thorpe [9] have shown that these values observe regularities as a function of their rank across natural images. This regularity is due to the regular distribution in natural images of singularities of different orders (i.e. in order: dots, lines, ramps, gradients) which are related to the wavelet coefficients (see Mallat [3, p. 513]) and which are ordered from the highest to the lowest in this algorithm. A solution is therefore to use the mean analog value to form a look-up table (LUT) to decode the analog values from their rank. However, we proved [6] that this regularity is enhanced if each scale is tuned so that the LUT corresponding to the different scales add up harmoniously according to the statistics of natural images [1]. This is done by tuning

[1] Instead of differentiating ON or OFF cells, we will consider for simplicity that each neuron i is assigned a polarity p_i which is either $+1$ or -1, so that the coefficients are rectified (i.e. $|C_i| = p_i \,.\, C_i$).

[2] This condition is approximately met in Van Rullen and Thorpe [9].

the norm N_i of the filters so that the coefficients appear with the same distribution at every scale. Finally, this proves that this strategy builds a complete and efficient code from the retina (analog to spike coding) which we may decode (spike to analog coding) using solely the rank of the spikes in the wave front, i.e. a *rank-order coding* scheme [7]. This strategy thus forms a *compact spike code* for static image representation.

2. Constructing lateral interactions

2.1. Orthogonal vs. non-orthogonal representations

The condition on the filters for a perfect reconstruction, i.e. the orthogonality of the dictionary used to represent the image—is a strong constraint on the architecture and is achieved only approximatively with the model presented in [9], resulting in a small information loss. Moreover, in the biological retina, the architecture is not dyadic and neighboring biological neurons most often have correlated sensitivities. The orthogonality condition is therefore too restrictive in order to build a biologically inspired model of the retina but also to apply the algorithm in further models in the primary visual system where the interdependence is even stronger.

More importantly, this representation is sensitive to usual transformations as small translations or rotations. In fact, in order to code the image in a more stable way, we may want to use an *over-complete representation* of the image, i.e. for which the number of filters is far greater than in the previous model. But applying that representation with a similar wavelet architecture would yield a highly redundant code, and we rather need that it defines a *sparse code*, i.e. that the model's coefficients absolute values rapidly decrease [5]. But mathematically optimizing the linear generative model leads to a combinatorial explosion of the freedom of choice of the filters and of their corresponding coefficient values (it is a *NP-hard* problem [3]).

2.2. Spike coding using a matching pursuit

Another strategy is to use a matching pursuit (MP) [3, pp. 412–419] algorithm, which is derived from a statistics' estimation algorithm and was also extended to wavelet theory [4]. The idea is that we have to account for the correlations between filters and we therefore need to build up lateral interactions to cancel the correlation whenever a filter is selected.

The MP algorithm decomposes the image over a dictionary $\mathcal{D}$ by iteratively choosing the best match and then—in order to minimize the residual energy knowing this match —removing the orthogonal projection of this match. Let us initially set the initial residual image $I^0 := I$ and activities $C_i^0 := C_i$ at the initial time $t = 0$. First, we determine the first neuron in the layer to fire as the most activated

$$i^0 = \mathrm{ArgMax}_i(|C_i^0|)$$

and for this neuron of index i^0, we define the corresponding extremal contrast value $C_{i^0}^0$. Actually, we found the best match in the sense of the projection of the image on the dictionary, so we subtract the projection of this match to I^0 in order to define a first residual I^1 at time $t=1$.

$$I^1 = I^0 - \frac{\langle I^0, \phi_{i^0}\rangle}{\|\phi_{i^0}\|^2} \cdot \phi_{i^0} = I^0 - \frac{C^0_{i^0}}{N^2_{i^0}} \cdot \phi_{i^0},$$

where N_{i^0} denotes the norm of filter i^0. Assuming the existence of fast interneuronal pathways, the activity may be directly updated at time $t=1$ from Eq. (1):

$$C^1_i = \langle I^1, \phi_i\rangle = C^0_i - \frac{C^0_{i^0}}{N^2_{i^0}} \cdot \langle \phi_{i^0}, \phi_i\rangle.$$

In particular $C^1_{i^0}=0$, i.e. the activity corresponding to the best match at time 0 is totally canceled at time 1. In neuronal terms we do not need to update the image's intensities (backward propagation) but directly the activities (lateral propagation). Iterating these steps, we may repeat this algorithm to obtain successive residuals at the discrete times t defined by the algorithm. This algorithm is exactly equivalent to MP for normalized filters ($N_i = 1$) and presents the same computational complexity and properties [3, pp. 412–419] and in particular the convergence of the reconstruction [3, p. 414].

As with the wavelet transform, it may be similarly translated to a spike coding scheme by associating to each step the firing of a spike, so that it is simply for $t \geqslant 0$,

$$\begin{cases} i^t = \mathrm{ArgMax}_{i\in\mathscr{D}}(|C^t_i|) \\ C^{t+1}_i = C^t_i - p^t \,.\, m^t \,.\, \langle \phi_{i^t}, \phi_i\rangle \end{cases}$$

with $m^t = |C^t_{i^t}|/N^2_{i^t}$ and p^t is the sign of $C^t_{i^t}$ (i.e. its ON or OFF polarity). We therefore associate to each spike a lateral interaction $\langle \phi_{i^t}, \phi_i\rangle$ which accounts for the selected spike. The reconstruction is then simply

$$I_{\mathrm{rec}}(T) = \sum_{t=0,\dots,T} p^t \,.\, m^t \,.\, \phi_{i^t}.$$

The choice of a match is fed back to the neurons' activities as a lateral interaction proportional to $p^t \,.\, m^t$ and to the correlation between the filters $\langle \phi_{i^t}, \phi_i\rangle$. With an over-complete dictionary, this coding strategy provides a *sparse representation* of the signal. In comparison with a wavelet decomposition, since the choice of the nth filter depends on the spike list for the previous times, this transform is non-linear.

2.3. Rank order coding with MP

To compare this algorithm with the model of Van Rullen and Thorpe [9], we kept the same architecture and observed the behavior of the absolute coefficients' values in function of the rank of propagation for different natural images drawn from a database of indoor and outdoor scenes. Similarly as the previous model, we observed regularities across natural images and that this behavior showed up to be stable allowing the similar

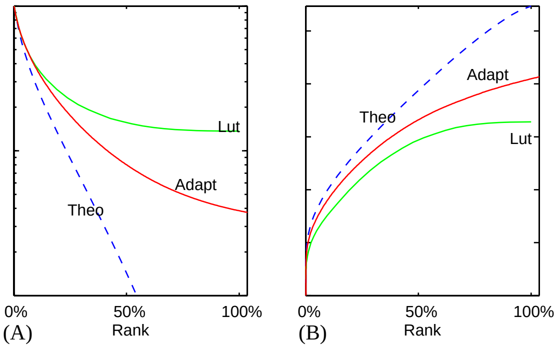

Fig. 1. Rank order coding with MP in the model retina. For the architecture defined in [9] we calculated (A) mean squared error and (B) mutual information of the reconstruction in function of the relative rank (the percentage of the number of spikes fired to the total number of neurons) for the different coding strategies, comparing (Theo) the theoretical reconstruction from the orthogonal wavelet coefficients, (Lut) the orthogonal wavelet coding using a LUT as in [9], and (Adapt) MP with online learning (the image database consisting of 100 images to learn the modulation function and 100 images to measure the reconstruction error). The adaptability of the MP algorithm enhances the transmission of the image and proves the possible use of the relative order of the action potentials as a code in the optic nerve.

use of a LUT to decode the analog value by its rank. Moreover, we used an incremental adaptive rule which has the advantage of being more biologically plausible and enabling on-line learning. This rule takes the form of a stochastic algorithm so that after coding the nth image using $m^{(n)}$ as a modulation function,

$$m^{(n+1)}(t) = (1 - \mu^{(n)}) \,.\, m^{(n)}(t) + \mu^{(n)} \,.\, |C^t_{i^t}|,$$

where t is as before the discrete time corresponding to the decomposition and $\mu^{(n)}$ (typically, $\mu^{(n)} = 1/n$) the stochastic learning gain. Practically it shows similar behavior as the LUT and leads to a similar reconstruction error.

Using the mean absolute coefficients as a LUT, we thus built a mechanism of reconstruction from the spike list, but as opposed to [9], this algorithm is adaptive and therefore the error may be compensated dynamically. Though filters are almost orthogonal (so that lateral interactions between filters—i.e. their correlation—is relatively low) the MP algorithm introduces a gain in the sparsity of the coefficients but also in the reconstruction quality (see Fig. 1).

3. Sparse spike coding

3.1. Extension to a multi-layer spike code

To model the multi-layered architecture of the visual system, we may easily extend this strategy to a multi-layer architecture. In fact, since the activity at the synapses of the neurons of the first layer may be incrementally constructed as $L(t+1) = L(t) + p^t \,.\, m^t \phi_{i^t}$, and the activity $C_{2,j}$ at the second layer consisting of the neurons j can directly be incrementally computed at retinal time t as

$$C_{2,j}(t+1) = \langle L(t+1), \psi_j \rangle = C_{2,j}(t) + p^t \,.\, m^t \,.\, \langle \phi_{i^t}, \psi_j \rangle,$$

where ψ_j is the weight vector of neuron j (e.g. orientation selective in V1) and so that we do not need to compute nor reconstruct the intermediate image $L(t)$. By defining an appropriate threshold for this activity, we may build a spiking mechanism and initiate lateral interactions similar to MP, so that at the t'th firing in V1 of a neuron $j^{t'}$, the activity is updated as

$$C_{2,j}(t') \leftarrow C_{2,j}(t') - C_{2,j'}(t') \,.\, \left\langle \frac{\psi_{j^{t'}}}{N^2_{2,j^{t'}}}, \psi_j \right\rangle,$$

where $N_{2,j} = \|\psi_j\|$. This scheme is still similar to MP (in particular $C'_{j'}(t'+1) = 0$) and its convergence theorem still holds.

We applied this algorithm with an over-complete dictionary, similar to the set of filters observed for the simple cells of the primary visual area, V1 (in humans the number of GCs is of the order of one million as for V1 the number it is approximately 300 million). Subsequently, we used the same method with a set of weight vectors ψ_j defined as dilated, translated and sampled DOG and Gabor filters (see [3, p. 160]). The scale grows geometrically with a factor $\rho = \sqrt[5]{2}$ (i.e. 5 layers per octave) on 41 scales and the direction is circularly NULL (i.e. a DOG), 0, $\pi/4$, $\pi/2$ and $3\pi/4$. The resulting distribution of the coefficients is highly kurtotic and the LUT were tabulated in the same manner,[3] so that the information rate—i.e. the information needed to code the address of one spike—is in this layer $\sim 16.1^{\text{bit}}$/spike. Convergence is quicker (see Fig. 2) so that this code may be compared to JPEG at high compression gains as we have shown in [6].

3.2. Propagation heuristic: dynamical saliency

The sparse spike coding algorithm introduces a computation by events, here the neurons' action potentials, which introduces new techniques and algorithms to dynamically process the data flow and which may be enlarged to other events as collective bursts of populations of neurons. Even for our simple test case, coding a static image, changing the priority of events may modify the progressive transmission of the different components of the image, hence its processing.

In particular, it is possible to modify the spatio-temporal spike pattern by modifying the sensibility of some neurons over others, that is here by modifying the norm of the neurons in space [8]. At first, by giving more weight to the attended part of the image, we force these neurons to fire first. We may extend this model of an attentive mechanism by selecting in an unsupervised manner the attended region in function of

[3] Regularity for these filters in natural images is observed if their mean spectral energy is rotation invariant. Systems for urban images which typically show more horizontal and vertical lines should be tuned accordingly.

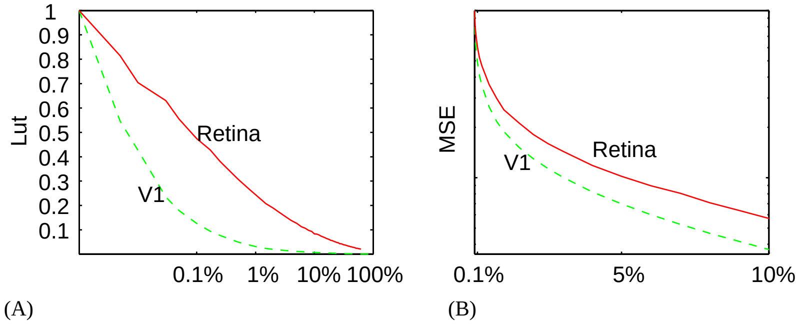

Fig. 2. Spike coding in the Retina and in V1. (A) We computed recursively the LUT for the model of the retina and for the model of V1 as a function of the rank (in percentage of the total number of neurons). In comparison with the retina, coefficients decrease more rapidly for the model of V1. (B) MSE for the corresponding progressive image reconstruction (using logarithmic *y*-axis) defined by using this spike code. The rapid convergence for V1 proves that we defined an efficient visual code using an over-complete set of Gabor filters and which leads to a model of a sparse spike code.

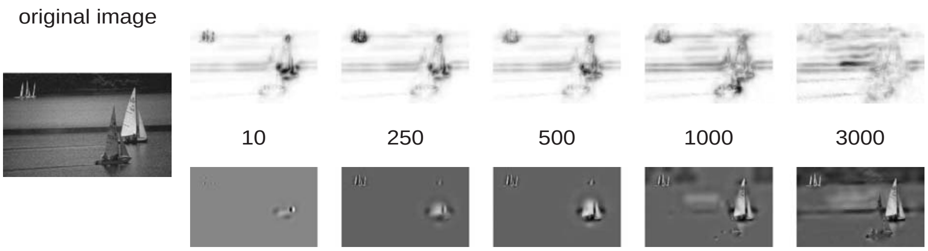

Fig. 3. Attentive mechanism. The MP algorithm, implementing the propagation in V1 of (Left) the ‘Sailboats’ image, was modified by a simple propagation heuristic favoring neurons in the vicinity of the previous firings. For illustration purpose we reconstructed (top row) a *dynamic saliency map* similar to [2] showing the coefficient’s energy image during the propagation (resp. from left to right for 10, 250, 500, 1000 and 3000 spikes). The maximum progressively shifts from one boat to the other and then to the background. The propagation is modified so that (bottom row) the reconstruction corresponds to a bottom-up attentive mechanism, revealing the corresponding objects. This propagation, compared to [2], permits the reconstruction of the image while using a much simpler architecture and avoiding an arbitrary definition of the inhibition of return by the implementation of the lateral interactions.

the firing history. For instance, at a step of the MP algorithm, instead of choosing the best match over all possible neurons, the match at time t should be chosen as

$$i^t = \mathrm{ArgMax}_{i \in \mathscr{D}}(|C_i^t| + \lambda \|i - i^{t-1}\|),$$

where λ is a regularization constant and $\|i - j\|$ is a distance measure between two neurons. Results show that this simple heuristics models a bottom-up model of visual attention (see Fig. 3, top row) by defining a dynamical saliency map which progressively reconstructs the image according to the attention areas (see Fig. 3, bottom row). This property may be useful for pattern recognition and especially when dealing with huge amount of data.

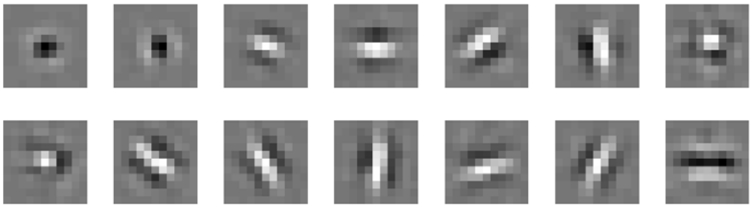

Fig. 4. Emergence of filters using an adaptive scheme. We simulated the propagation of the visual information through the retina and then through a second layer the filters of which are initially randomly set. By reinforcing the receptive fields of neurons toward the patch that elicited the firing of a neuron, we progressively extract primitives of the image which compete through the MP algorithm. Using the same protocol as Olshausen and Field [5], V1-like orientation selective filters similarly emerge, but the parallel propagation avoids the formation of doubles. These filters are therefore naturally centered and distinct.

3.3. Adaptive schemes

Following the analogy of MP with *vector quantization*, we may also construct a simple adaptive scheme for the filters using a modified Lloyd algorithm. We used the same protocol as Olshausen and Field [5] in a multiscale architecture (using a Gaussian pyramid) and progressively learned at each spike in this layer the filters toward the patches $I_{\mathscr{R}_{t'}}$ in the image that elicited the response of neuron $j^{t'}$ at time t'.

$$\psi_{j^{t'}} \leftarrow \psi_{j^{t'}} + \gamma \,.\, C'_{j^{t'}}(t') \,.\, I_{\mathscr{R}_{t'}},$$

where γ is the learning factor. We compared this algorithm with the Sparsenet scheme [5]. Even if similar in spirit (looking for a dictionary allowing for sparse coding), these algorithms differ in the sense that Sparsenet is biologically unrealistic but uses an efficient analytical optimization algorithm. Furthermore, it works on random image patches whereas we do not choose a priori a patch in the image but we rather progressively decompose the image in a *parallel* and *asynchronous* competition (see Fig. 4).

4. Conclusion

We have shown that we may define a complete spike code based on an arbitrary over-complete dictionary using lateral interactions defined by a MP algorithm and that this code is both efficient and sparse as is observed in the primary visual system. Experiments show the importance of the statistics of natural images as a chief constraint on the tuning of this algorithm to achieve good and sparse representation.

The sparse spike code algorithm leads to a general adaptive model of cortical processing which leads to a model of an ensemble of cortical neurons, the *cortical column*, as an autonomous multi-state automaton with strong interactions with neighbors and associated areas.

Finally, these algorithms advocate for a dynamical model of visual processing and provide a necessary extension of the rank order coding scheme by providing an efficient representation for visual tasks. This could provide efficient real-time applications using artificial asynchronous neural network which could mimic nature's performance.

References

[1] J.J. Atick, A.N. Redlich, What does the retina know about natural scenes? Neural Comput. 4 (2) (1992) 196–210.

[2] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259, URL: citeseer.nj.nec.com/itti98model.html.

[3] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, New York, 1998.

[4] S. Mallat, Z. Zhang, Matching pursuit with time–frequency dictionaries, IEEE Trans. Signal Process. 41 (12) (1993) 3397–3414.

[5] B. Olshausen, D.J. Field, Sparse coding with an over-complete basis set: a strategy employed by V1? Vision Res. 37 (1998) 3311–3325.

[6] L. Perrinet, M. Samuelides, S. Thorpe, Coding static natural images using spiking event times: do neurons cooperate?, IEEE Trans. Neural Networks, Special issue on Temporal Coding for Neural Information Processing (2004), URL: laurent.perrinet.free.fr/publi/perrinet03ieee.pdf

[7] S.J. Thorpe, D. Fize, C. Marlot, Speed of processing in the human visual system, Nature 381 (1996) 520–522.

[8] R. Van Rullen, S.J. Thorpe, Spatial attention in asynchronous neural networks, Neurocomputing 26–27 (1999) 911–918.

[9] R. Van Rullen, S.J. Thorpe, Rate coding versus temporal order coding: what the retina ganglion cells tell the visual cortex, Neural Comput. 13 (6) (2001) 1255–1283.