J Comput Neurosci (2012) 32:425–441
DOI 10.1007/s10827-011-0361-9

Relative spike time coding and STDP-based orientation selectivity in the early visual system in natural continuous and saccadic vision: a computational model

Timothée Masquelier

Received: 24 March 2011 / Revised: 5 September 2011 / Accepted: 8 September 2011 / Published online: 21 September 2011

Abstract We have built a phenomenological spiking model of the cat early visual system comprising the retina, the Lateral Geniculate Nucleus (LGN) and V1's layer 4, and established four main results (1) When exposed to videos that reproduce with high fidelity what a cat experiences under natural conditions, adjacent Retinal Ganglion Cells (RGCs) have spike-time correlations at a short timescale (~30 ms), despite neuronal noise and possible jitter accumulation. (2) In accordance with recent experimental findings, the LGN filters out some noise. It thus increases the spike reliability and temporal precision, the sparsity, and, importantly, further decreases down to ~15 ms adjacent cells' correlation timescale. (3) Downstream simple cells in V1's layer 4, if equipped with Spike Timing-Dependent Plasticity (STDP), may detect these fine-scale cross-correlations, and thus connect principally to ON- and OFF-centre cells with Receptive Fields (RF) aligned in the visual space, and thereby become orientation selective, in accordance with Hubel and Wiesel (Journal of Physiology 160:106–154, 1962) classic model. Up to this point we dealt with continuous vision, and there was no absolute time reference such as a stimulus onset, yet information was encoded and decoded in the *relative* spike times. (4) We then simulated saccades to a static image and benchmarked relative spike time coding and time-to-first spike coding w.r.t. to saccade landing in the context of orientation representation. In both the retina and the LGN, relative spike times are more precise, less affected by pre-landing history and global contrast than absolute ones, and lead to robust contrast invariant orientation representations in V1.

Keywords Early visual system · Continuous vision · Saccades · Spike time correlations · STDP · Neural coding

Abbreviations

CGC	Contrast Gain Control
DoG	Difference-of-Gaussian
EPSP	Excitatory Post-Synaptic Potential
IPL	Inner Plexiform Layer
IPSP	Inhibitory Post-Synaptic Potential
LGN	Lateral Geniculate Nucleus
LTD	Long Term Depression
LTP	Long Term Potentiation
OPL	Outer Plexiform Layer
PSTH	Post-Stimulus Time Histogram
RF	Receptive Field
RGC	Retinal Ganglion Cell
SRM	Spike Response Model
STDP	Spike Timing-Dependent Plasticity
V1	primary visual cortex (a.k.a. area 17).

1 Introduction

Vision is an ongoing process. The retina constantly performs a spatiotemporal filtering of the optical signal and encodes the result in spikes. These spikes are

Action Editor: Ken Miller

Electronic supplementary material The online version of this article (doi:10.1007/s10827-011-0361-9) contains supplementary material, which is available to authorized users.

T. Masquelier (✉)
Unit for Brain and Cognition,
Department of Information and Communication Technologies,
Universitat Pompeu Fabra,
Barcelona, Spain
e-mail: timothee.masquelier@alum.mit.edu
URL: http://alum.mit.edu/www/timothee.masquelier

subsequently processed by the visual system as they flow in. Even the processing of still images is dynamical: either they suddenly appear at a given time from the dark (a paradigm extensively studied in the lab, but rather unnatural, that we call the "stimulus onset paradigm"), or they stimulate the retina after a body, head or eye movement. Furthermore, even when fixating, our eyes are continuously making microsaccades (Martinez-Conde et al. 2004). If both saccades and microsaccades are blocked, visual percepts fade very quickly, sometimes in 80 ms only (Coppola and Purves 1996). There is no such thing as steady visual processing.

Despite these well-known facts, the vast majority of visual system models deal with still images and make the implicit assumption of a steady regime (e.g. (Fukushima 1980; LeCun and Bengio 1998; Riesenhuber and Poggio 1999; Ullman et al. 2002; Serre et al. 2007)). We now refer to these as "steady models". The entry layer spatially filters the input image (usually with Difference-of-Gaussian (DoG) filters if mimicking the retina, or Gabor filters if mimicking V1), and the result is passed to subsequent layers for further processing.

Two rather different lines of research have brought time into those models. The first one are the so called "trace rule" models (Földiák 1991; Wallis and Rolls 1997; Rolls and Milward 2000; Stringer and Rolls 2000; Einhäuser et al. 2002; Spratling 2005; Masquelier et al. 2007), able to learn invariance (for e.g. to position and scale) from a series of frames in which objects undergo smooth transformations. The effects of such rules have been formalized in the so called Slow Feature Analysis theory (Wiskott and Sejnowski 2002). However these models still process the frames one by one, and usually only the synaptic modification—not the activity—depends on more than one frame. Furthermore, the timescales involved, in the order of the second, are much slower than the ones in which we are interested here.

The second line of research is the one of Thorpe and colleagues, who are interested in much faster timescales—down to the individual spikes—and model the transient activity in stimulus onset paradigms (VanRullen et al. 1998; Delorme et al. 2001; Delorme and Thorpe 2001; VanRullen and Thorpe 2001; VanRullen and Thorpe 2002; Guyonneau et al. 2005; Masquelier and Thorpe 2007). In these models, the entry layers spatially filter the input image and perform an "intensity-to-latency conversion": the cells, initially in a resting state, integrate the activation currents, so the stronger a unit is activated—for e.g. by the presence of a salient edge in its RF—the earlier it fires a first spike with respect to the stimulus onset. Those spikes are then propagated asynchronously throughout the hierarchy. These models constitute a plausibility proof for the use of time-to-first spike coding in stimulus onset paradigms, and may account for the reported phenomenal speed of visual categorization with these paradigms, estimated from both behavioural responses(Thorpe et al. 1996; Fabre-Thorpe et al. 1998; Rousselet et al. 2002; Bacon-Mace et al. 2005; Kirchner and Thorpe 2006; Serre et al. 2007; Girard et al. 2008; Crouzet et al. 2010) and electrophysiological recordings (Oram and Perrett 1992; Hung et al. 2005; Liu et al. 2009).

However, as mentioned above, stimulus onset paradigms are rather unnatural. A more natural situation is that an image is formed on the retina at $t = t_0$ after a body or head movement, a saccade, or a microsaccade (all of these are referred to as "movements" below). In that case the "intensity-to-latency conversion" hypothesis is questionable for several reasons. First, the input current to a RGC is a spatiotemporally filtered version of the luminance signal, as opposed to a mere spatially filtered version (among other things the surround signal is delayed, as we will see below). This spatiotemporal filtering does not stop during the movements. This means that the RGC input currents at $t = t_{0,}$ and slightly after, depend not only on the current image, but also on what happened during the movement, and possibly even before. Furthermore, these currents are integrated and converted into spikes, which introduces another dependence on history (the same input current may lead to different spike latencies, depending on when the last spike was emitted). For all these reasons the times-to-first-spikes with respect to t_0 are expected to be poor encoders of the current image.

Here we claim that the *relative* spike times encode the images more robustly, because (1) history typically influences neighbouring cells' spike times similarly, and thus has a weak effect on their relative spike times (2) input currents produced by natural videos are somewhat "sparse", that is sub-threshold most of the time (periods during which the history is quickly forgotten), but with rare high yet short supra-threshold peaks. Thus spike time jitter cannot accumulate much, and the natural images' salient edges generate nearly synchronous spikes in neighbouring RGCs and LGN cells. Relative spike time coding has another advantage: it does not need the reference time t_0, therefore it can be at work in continuous, non-saccadic, vision. Relevantly, it is those relative spike times—not the absolute ones—that matter for downstream neurons in V1 and for STDP. These V1 neurons can thus progressively become orientation selective, even in continuous vision.

To demonstrate these points, we built a phenomenological spiking neuronal model of the cat early visual system's X feedforward pathway (see Fig. 1). This pathway is involved in fine vision and object recognition. We exposed the model to videos that approximate the natural input to the cat visual system (Betsch et al. 2004). A recent detailed retinal model was used to convert those videos into spikes (Wohrer and Kornprobst 2009), that were then asynchronously propagated in a feedforward network comprising a LGN stage and a V1 stage. We first observed that adjacent

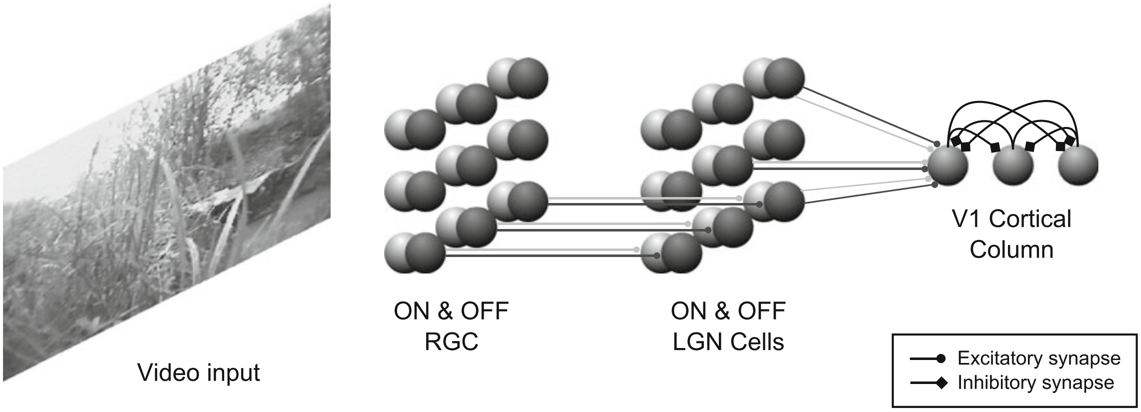

Fig. 1 Overview of the feedforward network. The video input is converted into spikes using the Virtual Retina simulator (Wohrer and Kornprobst 2009), which modeled both ON and OFF X RGCs, organized in 11×11 retinotopic maps (for clarity, we only represented 3×3×2 cells). Each ON (respectively OFF) RGC projects to one ON (resp. OFF) relay cell in the LGN, which is still retinotopically organized (for clarity we only represented 6 of the 11×11×2 connections). Each of the 32 downstream V1 simple cell in the cortical column (we only represented 3) integrates spikes from the same 11×11 ON and 11×11 OFF LGN cells through plastic synapses governed by STDP, and will thus extract visual patterns that are consistently present in the input (for clarity we only represented 6 of the 11×11×2×32 plastic synapses). Lateral inhibitory connections between those cells prevent them from learning the same patterns (the apparent violation of Dale's law can be resolved via inhibitory interneurons)

RGCs with same polarity have spike-time correlations at a short timescale (~30 ms). The cells thus tend to fire simultaneously (each time a salient edge enters their receptive fields) despite having a different history and receiving independent input noise. Then we showed that the LGN filters out some noise and thus increases the spike reliability and temporal precision, the sparsity, and, importantly, further reduces the timescale of adjacent cells' spike-time correlation down to ~15 ms. It thereby facilitates the STDP-based learning of downstream V1 simple cells, which gradually became orientation selective, by connecting preferentially to LGN cells with RFs aligned in the visual space. All of this happened without absolute reference times such as saccade or stimulus onsets, i.e. in a clock-free system. We then simulated saccades, whose landing times provided reference times. But, as expected, it turned out that the times-to-first-spikes with respect to those landing times encoded the target image orientations poorly. Relative spike times did a better job at it.

2 The phenomenological model

Here we describe in detail the stimuli we used, and our feedforward model of the cat X-pathway (Fig. 1). All the code has been made available:

- The Virtual Retina simulator is available here: http://www-sop.inria.fr/neuromathcomp/public/software/virtualretina/
- The custom Matlab/C code for LGN and V1 stages is available here: http://senselab.med.yale.edu/modeldb/ShowModel.asp?model=141062

2.1 Retina

We used the Virtual Retina simulator (Wohrer and Kornprobst 2009), a detailed retina model with contrast gain control (CGC), which reproduces many experimental findings. The simulator is highly configurable through a xml parameter file, and we tuned it to mimic the cat X RGCs in the fovea region. Table 1 gathers all the numerical parameters, and the corresponding sources. The reader should refer to (Wohrer and Kornprobst 2009) for further information.

One important consequence of the delayed surround (τ_S=5 ms) is that the resulting filter is spatiotemporal and not-separable in time and space (Wohrer and Kornprobst 2009), and not a simple DoG spatial filter, as often assumed in visual models.

We used maps of 11×11 ON- and 11×11 OFF-RGCs, each one projecting to one LGN cell (see Fig. 1).

2.2 Stimuli

The videos we used were collected by König's group (Betsch et al. 2004). They mounted cameras to cats' heads, and recorded while the animals were free to explore a natural environment. These videos thus approximate the input to which a cat visual system is naturally exposed, preserving its spatiotemporal structure. Eye movements, however, are not taken into account, but their effect is supposed to be negligible since freely behaving cats make mostly eye movements of small amplitude (Betsch et al. 2004). The camera spans a visual angle of 71° by 53° and its resolution is 320×240 pixels. The original sampling rate was 25 frames per second. For reasons explained in the

Table 1 Retina parameters

Parameter	Value	Comment	Source
Outer Plexiform Layer (OPL)			
σ_C	0.15°	Centre gaussian's sigma	(Wohrer and Kornprobst 2009)
τ_C	10 ms	Centre signal low pass filtering time constant.	(Wohrer and Kornprobst 2009)
τ_U	100 ms	Undershoot high pass filtering time constant.	(Wohrer and Kornprobst 2009)
w_U	0.8	Undershoot transient relative weight.	(Wohrer and Kornprobst 2009)
σ_S	0.8°	Surround gaussian's sigma	(Wohrer and Kornprobst 2009)
τ_S	5 ms	Surround signal low pass filtering time constant.	(Enroth-Cugell et al. 1983; Cai et al. 1997)
λ_{OPL}	10 Hz/Lum. unit	Overall gain of the centre-surround filter.	(Wohrer and Kornprobst 2009)
w_{OPL}	1	Relative weight of centre and surround signal.	(Wohrer and Kornprobst 2009)
Use leaky heat equation	True	Averaging by gap junctions rather than dendritic spread. Leads to a non-separable spatio-temporal filter, but somewhat more realistic.	(Wohrer and Kornprobst 2009)
Contrast Gain Control (CGC)			
λ'_{OPL}	20	Another gain applied right after λ_{OPL}, thus without biological meaning, but useful for implementation issues.	(Wohrer 2008)
g_A^0	5 Hz	Inert leaks in membrane integration.	(Wohrer and Kornprobst 2009)
σ_A	1°	Size of the spatial neighbourhood used to estimate local contrast.	(Wohrer and Kornprobst 2009)
τ_A	5 ms	Size of the temporal neighbourhood used to estimate local contrast.	(Wohrer and Kornprobst 2009)
λ_A	50 Hz	Strength of the gain control feedback loop.	(Wohrer and Kornprobst 2009)
Inner Plexiform Layer (IPL)			
τ_G	20 ms	High pass filtering time constant.	(Wohrer and Kornprobst 2009)
w_G	0.7	Transient relative weight.	(Wohrer and Kornprobst 2009)
σ_G	0°	No additional pooling for X cells.	(Wohrer and Kornprobst 2009)
v_G^0	0	Bipolar linear threshold.	(Wohrer and Kornprobst 2009)
λ_G	150 Hz	Slope in the linear area.	(Wohrer and Kornprobst 2009)
i_G^0	37 Hz	This is below the threshold current (50 Hz). Thus in the dark the threshold is reached only because of the noise (see below), which leads to a irregular Poisson-like spontaneous activity (at ~1 Hz), in accordance with experimentation in cats:	(Kara et al. 2000)
Retinal Ganglion Cells (RGC)			
g^L	50 Hz	Leak conductance (thus the membrane time constant is 20 ms)	(Wohrer and Kornprobst 2009)
σ_v	0.1	Gaussian white noise current's normalized amplitude (see Eq. (1)). Integration of this current leads to a Gaussian auto-correlated process with time constant $1/g^L$ and variance σ_v. The numerical value has been estimated experimentally in cats:	(Keat et al. 2001)
η_{refr}	N(3 ms,1 ms)	Refractory period is normally distributed with mean 3 ms and standard deviation 1 ms	(Wohrer and Kornprobst 2009)
Density	17 cells/°	RGC density in the cat fovea region is about 6,000/mm^2, and the focal length is about 12.5 mm, leading to ~17 cells/° (that is a mean inter-RGC interval of 0.06°)	(Stone 1965)

next section this sampling rate was too slow for our study, and we interpolated the frames to 100 frames/s using the VirtualDub freeware (http://www.virtualdub.org). In the low contrast condition of Section 3.4 we decreased the RMS contrast to 50% of its original value, while the mean luminance was unchanged. This was done by linearly rescaling the distances between individual pixel luminances and the mean one. No other pre-processing was done. Unless said otherwise, we only used the first movie out of the 17 they recorded.

2.3 Interpolation avoids RGC phase locking

As said above, the original videos were recorded at 25 frames/second (Betsch et al. 2004). Is this sampling rate sufficient to capture the dynamics of the world from a cat's perspective? It does not seem so: this world turns out to contain a significant amount of energy in frequency bands above 12.5 Hz (see Fig. 2a, light grey line), probably due the cats' rapid head movements. These high frequencies are lost in the discretized signal (Nyquist theorem), but this creates discretization artifacts, that is some pixel values change abruptly from one frame to the next one. As a result, when feeding these frames to the retina simulator at the rate of 25 frames per second, the RGC input current power spectrum showed a peak at 25 Hz (Fig. 2a), despite low pass filtering in the OPL and bipolar cells (see (Wohrer and Kornprobst 2009) for details). The RGC unconstrained potentials, that is before thresholding, result from the leaky-integration of those currents, plus a Gaussian white noise $\xi(t)$(with $\langle\xi(t)\rangle = 0$ and $\langle\xi(t)\xi(s)\rangle = \delta(t-s)$), leading to the Langevin equation:

$$\frac{dV}{dt} = g^L(RI - V) + \sigma_v\sqrt{2g^L}\xi(t) \tag{1}$$

This leaky integration further low pass filters the signal with a cut off frequency of $g^L/2\pi \approx 8$Hz. Yet the membrane potential's power spectrum still has a significant peak at 25 Hz even when a realistic level of noise—$\sigma_v = 0.1$ (Keat et al. 2001)—was used (see Fig. 2a, top). As a result, the RGC spikes tend to phase-lock to the frame onset (see Fig. 2a, bottom). This phenomenon is probably not an artifact of the model: likely, it would also happen if the cat was watching the video. However, it would not happen it the cat was watching the real, continuous world—the situation in which we are interested here.

This is the reason why we interpolated the frames to 100 Hz (see Stimuli section above) (50 Hz turned out to be insufficient). A small peak was still visible at 100 Hz in the power spectrum of RGC current and potential without noise $\sigma_v = 0$ (Fig. 2b, top), but it disappeared in both when adding noise. Consequently, the RGC spikes were not

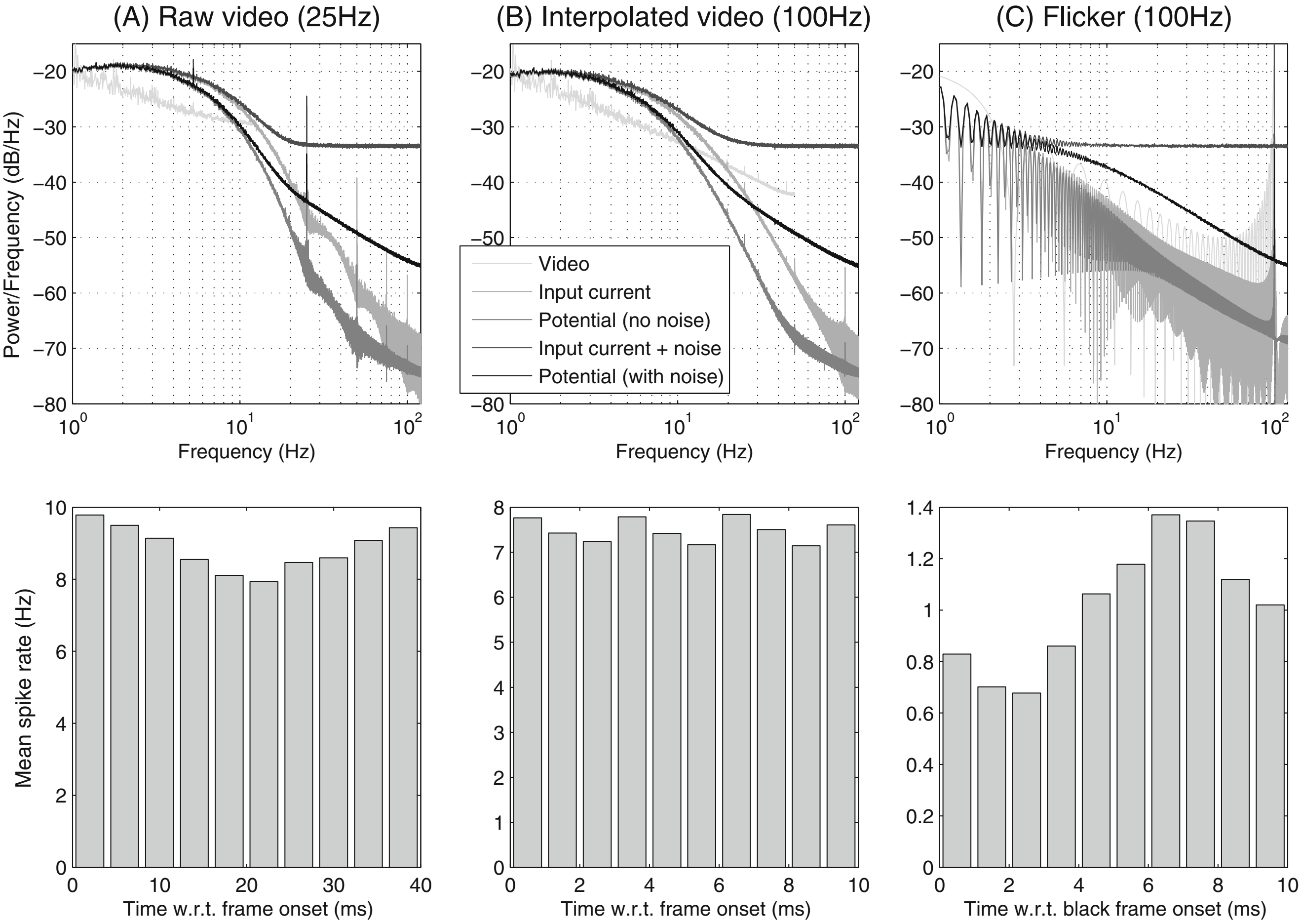

Fig. 2 Power spectrums and phase locking. Top row shows the power spectrum of the video signal, RGC input current and unconstrained potential, with and without noise (estimated with periodograms). For normalization issues the video raw luminance values were divided by 128. Bottom row shows histograms of RGC spike times w.r.t to frame onsets. See text for details.

entrained by the frame rate anymore (Fig. 2b, bottom). Also notice the peak at 25 Hz in the no-noise case. This is because interpolated frames tend to be more blurred than the original ones, so one every four frames has more saliencies. However this peak also disappears when adding noise, and we checked that the spikes did not lock to this 25 Hz rate either. We thus assumed that the interpolated video at 100 Hz and the continuous world were equivalent from the retina's point of view, and we used this interpolated video in the rest of the study.

This 100 Hz rate seems sufficient to sample the world from a cat's perspective, due to its frequency content, generating a smooth frame sequence, to which the RGC spikes do not lock. However, this does not mean that 100 Hz-and-above frequencies are filtered out by the retina. To make this point clear, we exposed the retina to a 100 Hz flicker, by alternating black and white frames at 200 Hz. The resulting video signal had a huge peak at 100 Hz, that propagated until the RGC potentials even in the noisy case (Fig. 2c top), and significantly entrain the spikes (Fig. 2c bottom). This is consistent with experiments in macaque and humans showing that even V1 neurons can be entrained by Cathode Ray Tube display refresh rates up to 100 Hz (Williams et al. 2004).

2.4 LGN

LGN relay cells were modelled with the Spike Response Model (SRM) (Gerstner et al. 1993). Each presynaptic spike j, with arrival time t_j, adds to the membrane potential an Excitatory Post-Synaptic Potential (EPSP) of the form:

$$\varepsilon(t-t_j) = K \cdot \left(\exp\left(-\frac{t-t_j}{\tau_m}\right) - \exp\left(-\frac{t-t_j}{\tau_s}\right) \right) \cdot \Theta(t-t_j) \tag{2}$$

where τ_m is the membrane time constant (here 20 ms), τ_s is the synapse time constant (here 5 ms),Θ is the Heavyside step function:

$$\Theta(s) = \begin{cases} 1 & if \quad s \geq 0 \\ 0 & if \quad s < 0 \end{cases} \tag{3}$$

and K is just a multiplicative constant chosen so that the maximum value of the kernel is 1 (the voltage scale is arbitrary).

Those EPSPs sum. At any time the membrane potential is thus:

$$p = \sum_{j/t_j > t_i} \varepsilon(t-t_j) \tag{4}$$

Notably, this EPSP summation at the retinogeniculate synapse is consistent with experimentation in macaques (Carandini et al. 2007). This SRM formulation allows us to use event-driven programming: we only compute the potential when a new presynaptic spike is integrated. We then estimate numerically if the corresponding EPSP will cause the threshold (1.25 for the LGN neurons) to be reached in the future and at what time. If it is the case, a postsynaptic spike is scheduled. Such postsynaptic spike events cause all the EPSPs to be flushed. There is then a refractory period of 3 ms, during which the neuron stops summing the EPSPs.

2.5 V1

The V1 neurons were modelled using the same algorithm as in (Masquelier et al. 2009a). We briefly describe it again here for the reader's convenience.

We used the SRM again but this time each V1 neuron integrates EPSP coming from the 11×11×2 LGN cells, through weighted synapses. Furthermore, when a neuron fires at time $t_k$it sends to all the others an Inhibitory Post-Synaptic Potential (IPSP). For simplicity we used the same kernel as for EPSP with a multiplicative constant α (here 0.5):

$$\mu(t-t_k) = -\alpha \cdot T \cdot \varepsilon(t-t_k) \tag{5}$$

At any time the membrane potential is thus:

$$p = \sum_{j/t_j > t_i} w_j \cdot \varepsilon(t-t_j) + \sum_{k/t_k > t_i} \mu(t-t_k) \tag{6}$$

where the w_j are the excitatory synaptic weights, between 0 and 1 (arbitrary units). Those weights are subject to STDP. We used a classic additive exponential update rule:

$$\Delta w_j = \begin{cases} a^+ \cdot \exp\left(\frac{t_j - t_i}{\tau^+}\right) & \text{if} \quad t_j \leq t_i \quad \text{(LTP)} \\ -a^- \cdot \exp\left(-\frac{t_j - t_i}{\tau^-}\right) & \text{if} \quad t_j > t_i \quad \text{(LTD)} \end{cases} \tag{7}$$

Following learning the weights were clipped to [0,1]. For each afferent, we also limited LTP (respectively LTD) to the last (first) presynaptic spike before (after) the postsynaptic one ('nearest spike' approximation).

We used $\tau^+ = 17$ and $\tau^- = 34$ms (which is in the range of experimental estimations (Caporale and Dan 2008)), a^+= 0.01 and a^-=0.0085.

For the neuronal parameters, we used again τ_m=20 ms, τ_s=5 ms, and a refractory period of 3 ms. The threshold, however, was set to 30.

3 Results

3.1 Cross-correlations between adjacent RGCs

The mean distance between two adjacent cat foveal RGCs is about 0.06° (Table 1), that is about 40% of their centre width σ_C. In other words, the RFs of adjacent RGCs significantly overlap, therefore the light they receive is correlated, and so is their membrane potential, as can be seen in Fig. 3 (top), which for clarity does not include the noise (σ_v=0). Also note an important property of the potential time courses: they are somewhat "sparse", that is weak most of the time, but with rare short and high peaks. It is thus possible to fix a threshold above which the unconstrained potential only makes short incursions (suprathreshold mode), producing typically 1–2 spikes per peak, and under which it remains most of the time (subthreshold mode). Why is that important? When in suprathreshold mode a neuron behaves as an integrator, and transmits temporal patterns with only low reliability (König et al. 1996). This is because the same input current pattern can lead to very different output spike times, depending on the initial condition on the membrane potential, which typically depends on when the last spike was emitted, and which is never forgotten. Thus spike time jitter can accumulate (Brette and Guigon 2003). Conversely, when in subthreshold mode the history is quickly forgotten (with time constant $1/g^L$= 20 ms here), and potential and unconstrained potential quickly join (Fig. 3a, solid and dashed curves). The net result is that adjacent RGCs tend to fire quasi-simultaneously (each time a salient edge enters their receptive fields). Adding the (independent) noise (Fig. 3b) jitters the spikes, and produce a few extra ones, but the tendency to fire simultaneously is still there (quantified below in Fig. 4).

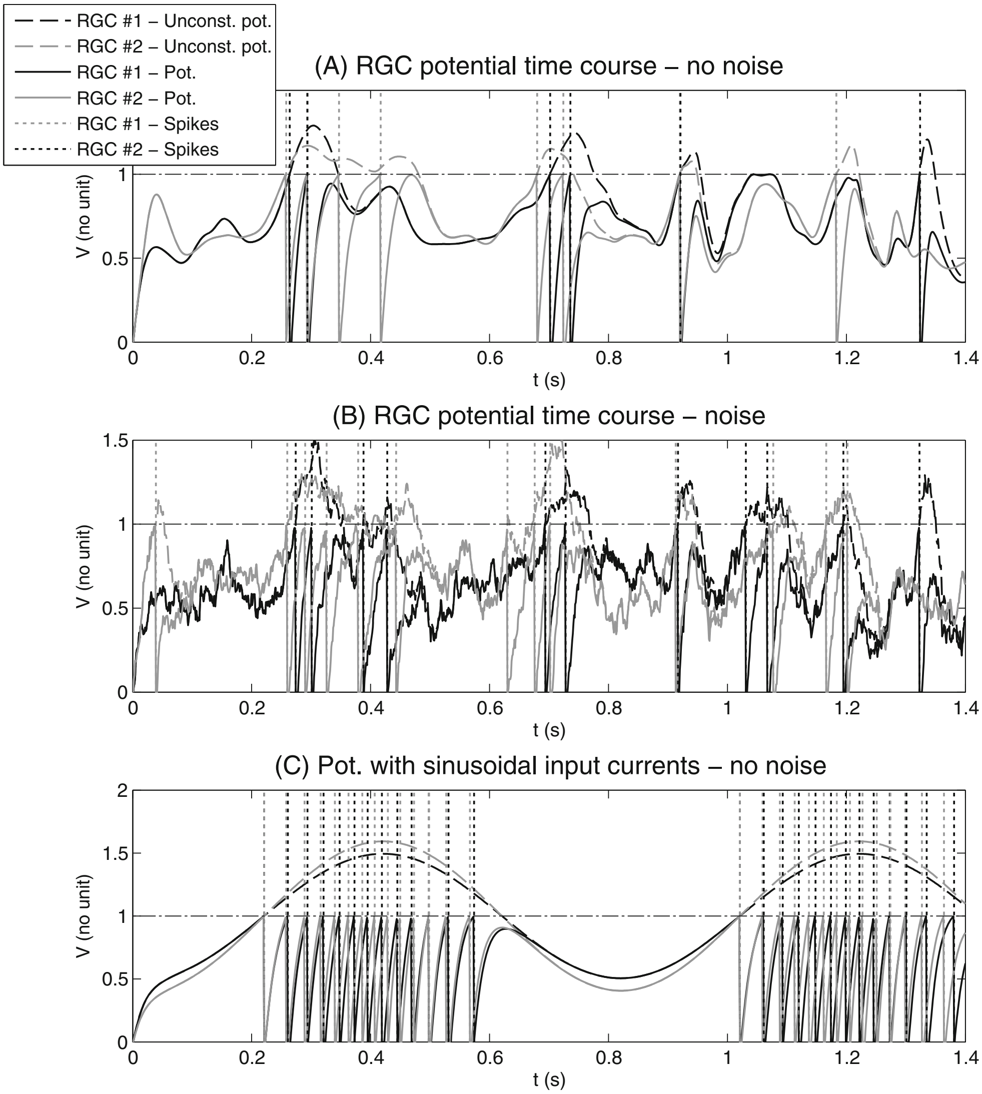

Fig. 3 Potential time courses and output spikes (*vertical doted lines*) for two adjacent RGCs (*one in black*, *one in grey*). The unconstrained potential ignores the threshold. (**a**) Without noise (**b**) With noise (**c**) Toy example to illustrate that correlated inputs does not necessarily imply synchronous output spikes (see text)

It is important to realize that the assumptions made so far are reasonable: we exposed a realistic retina model, whose parameters are in the biological ranges (Table 1), to realistic natural stimuli, and that led to correlations in adjacent RGC inputs that were transmitted in their output spikes (quantified below in Fig. 4c). This transmission is not automatic. As a counter example, consider two inputs slowly oscillating in phase (here at 1.25 Hz), thus perfectly correlated, with the same mean, but slightly different amplitudes (Fig. 3c). In this case the neurons stay too long in suprathreshold mode, when the grey neuron's inter spike intervals are lower than the black neuron's. These time differences accumulate, leading to a flat probability for the time lag between grey and black spikes. Thus the correlation in the inputs is not transmitted in the output spike times. This will happen as long as the current is suprathreshold for a period >> $1/g^L$, that is for f << 50 Hz.

Fortunately, natural signals have a different temporal structure, with short and high peaks, favourable for temporal pattern transmission. To quantify this, we computed the correlations in RGCs' unconstrained potential and output spikes (Fig. 4abc). Notice that:

1. The noise decreases all the correlation coefficients (Fig. 4abc, compare grey and black curves)
2. As far as the unconstrained potential is concerned (Fig. 4a):

   a. Autocorrelation decreases much faster than the one of the raw video luminance signal. This is because the retina has high pass filter stages (undershoot in

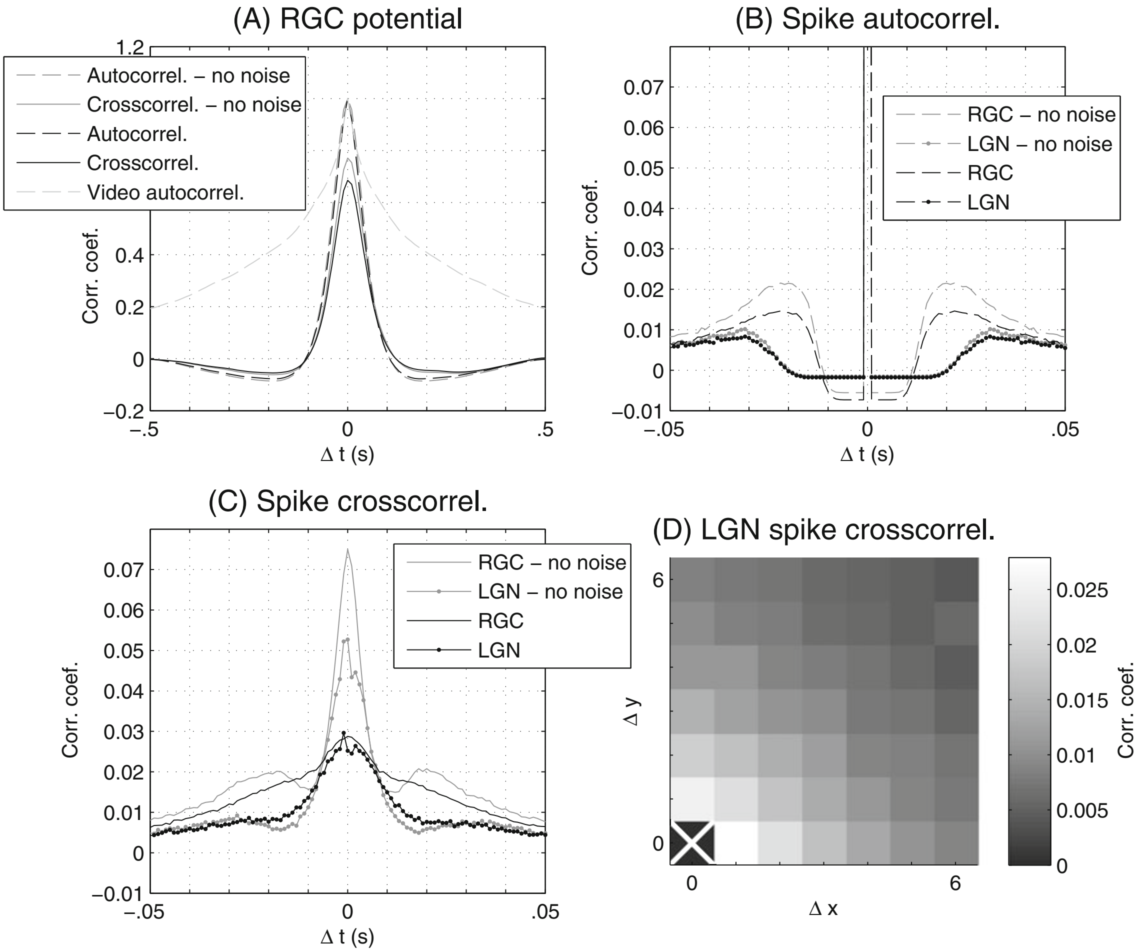

Fig. 4 Pearson correlation coefficient ($\in [-1, 1]$) in RGC/LGN inputs and outputs. Grey = without noise, black = with noise, dashed = autocorrelation, solid = crosscorrelation between adjacent cells, point markers = LGN, no marker = RGC (**a**) We plotted the correlation coefficient between a RGC unconstrained potential, and a time-shifted version of the same signal ("Autocorrel.") or of adjacent cell's unconstrained potential ("Crosscorrel."), with and without noise. For comparison, we also plotted the autocorrelation of the raw video luminance signal. (**b**) Correlation coefficient between the spike times (binned in 1 ms bins), and a time-shifted version of those spike times, with and without noise, for RGCs and LGN cells (**c**) Idem but the correlation is now computed between adjacent cells (**d**) Zero time-lag crosscorrelation coefficient between LGN spike times as a function of the distance in neurons (inter neuron distance=0.06°), with noise. For clarity issue we did not represent the autocorrelation case (*bottom left corner*), which would lead to a correlation coefficient of 1

the OPL, and phasic RGCs, see (Wohrer and Kornprobst 2009) for details).

b. The crosscorrelation between adjacent RGCs is about 2/3 of the autocorrelation

3. As far as the spikes are concerned (Fig. 4bc):

   a. Auto- and cross-correlations have the same order of magnitude (compare Fig. 4b and c), except for short time lags where the effect of the refractory period can be seen for the autocorrelation (Fig. 4b).

   b. Most importantly, the spike crosscorrelogram between two adjacent RGCs shows a significant centred peak (Fig. 4c, solid black line). This may appear obvious, but it was worth checking, because:

      i. Adjacent RGCs have shifted RF, so they receive a different (yet correlated) input.

      ii. Jitter accumulation might have prevented correlation transmission (see Fig. 3c).

      iii. The noise, independent for the two RGCs, might have flattened/broadened the peak completely.

As proposed in (Butts et al. 2007) we can fit the peak's top portion with a Gaussian, and the time lag at which the peak is 1/e time its height gives the crosscorrelation timescale, here 33 ms (the average jitter on individual spikes, if assumed to be Gaussian, is half this timescale).

4. We applied the same method to the RGC potential cross-correlations (Fig. 4a), and found a longer timescale of 55 ms. So only fast timescale correlations are transmitted from the potential to the spikes. This is consistent with Fig. 3c: slow potential correlations lead to flat spike cross-correlogram.

3.2 The LGN filters out noise

Real LGN relay cells receive feedforward input from only one to five RGCs (Rathbun et al. 2010). We restricted the model to monosynaptically connected LGN cells, organized in two retinotopic layers (ON and OFF) of 11×11 neurons (see Fig. 1). We used a threshold of 1.25, meaning that at that least two input spikes with a short inter-spike interval (15–20 ms) are needed to reach the threshold, in line with experimentation (Rathbun et al. 2010). Although quantitative comparison with (Rathbun et al. 2010) is not possible since they used white noise stimuli (as opposed to natural videos), our model qualitatively reproduce their findings:

1. Not all the RGC spikes are relayed by LGN cells, thus sparsity increases (here the average firing rate is 2 Hz in the LGN, versus 7 Hz in the retina).
2. Because noise mostly generates spikes isolated in time, these are not often relayed.
3. Relayed spikes are thus more reliable in their number, and more precise in their timing.

As a result, the crosscorrelation timescale for adjacent LGN cells (Fig. 4c, solid black line with point markers), is now of only 14 ms (against 33 ms in the retina). Relevantly, these crosscorrelation timescales have been estimated experimentally in cat LGN using the same video stimuli, and a mean value of ~15 ms was found (Desbordes et al. 2008), which is very comparable to our 14 ms. By narrowing down the crosscorrelation timescales, the LGN facilitates STDP-based learning by downstream V1 neuron (see next section). Not surprisingly, the crosscorrelation peak's height decreases with the distance between the cells (Fig. 4d).

In this study, we are more interested in cross- than auto-correlations. Yet in order to further validate quantitalively our model, we computed the auto-correlation timescale as well, also called "response temporal precision", using the same method as Butts and colleagues, who recorded in cat LGN using the same video stimuli (Butts et al. 2007). Specifically, we computed the PostStimulus Time Histogram (PSTH) of a single LGN X-cell, from 2,000 trials in which the first 10s of the video was presented. This PSTH's autocorrelation has a centred peak, and the time lag at which the peak is 1/e time its height gives the response timescale. Like them, we found a value of ~10 ms. This temporal precision is remarkable given that the natural stimuli tend to vary on a timescale that is several times slower (Butts et al. 2007).

We also computed the autocorrelation timescale in the retina using the same method, and found 23 ms. The LGN thus reduces both auto- and cross-correlation timescales. Table 2 gathers these results. In both the LGN and the retina, about 2/3 of the crosscorrelation timescale can be attributed to the noise, and the remaining third to the fact that adjacent cells receive slightly different inputs, because their RFs are slightly shifted.

We also computed the Fano Factor over the 2,000 trials (Table 2). It increases from the retina to the LGN, in line with experimentation (Kara et al. 2000; Rathbun et al. 2010) (this may seem to contradict our claim that the LGN filters out noise, but note that the spike count variance,

Table 2 Precision and reliability in the Retina and LGN

	RGC	LGN
Mean firing rate	7 Hz	2 Hz
Autocorrelation timescale	23 ms	10 ms
Crosscorrelation timescale	33 ms	14 ms
Fano factor	0.1	0.2

product of the mean spike count times the Fano Factor, decreases).

To visualize RGC and LGN spikes we used the jAER Open Source Project code freely available at http://sourceforge.net/apps/trac/jaer. This code allows building a video from a continuous spike flow. Specifically it constructed one frame every 20 ms, and represented each ON-cells' spike (respectively OFF-cells') on that time window by a white (resp. black) pixel (irrespective of the exact spike time within the 20 ms window). The results for 10s of simulated time, together with the video input, can be seen on the video provided in the Online Resource 1. Figure 5 shows one sample frame. Notice how the LGN filters out retinal noise.

3.3 STDP-based emergence of orientation selectivity in V1

According to the classic model proposed by Hubel and Wiesel (1962), simple cells in layer 4 of cat V1 (also known as area 17), receive primarily feedforward input from the LGN, and gain orientation selectivity by integrating inputs from LGN cells with RFs that are aligned in visual space—a proposition supported by direct physiological evidence in cats (Ferster et al. 1996; Chung and Ferster 1998) and ferrets (Chapman et al. 1991). Besides, evidence for STDP abounds in the visual cortex (see (Caporale and Dan 2008) for a review), and we propose here that the rule could account for the selection of aligned input. Specifically, we consider one V1 cortical column in which 32 neurons have feedforward plastic synapses with the same 11×11×2 LGN cells (see Fig. 1), leading to RF sizes of about 0.8°×0.8°, which matches well experimental estimations for simple cells in cat foveal V1 (Wilson and Sherman 1976). Lateral non-plastic inhibitory connections are set up between them, so that as soon as a neuron fires, it sends a strong inhibitory postsynaptic potential to its 31 neighbours (the apparent violation of Dale's law can be resolved via inhibitory interneurons). This tends to prevent the neurons from learning the same patterns (Masquelier et al. 2009a). Starting from random feedforward synaptic weights, we train the network with the natural videos, and Fig. 6 illustrates how orientation selectivity gradually emerges for most of the 32 neurons. Note that we approximated the cells' preferred stimuli by linear summation of simple DoG spatial filters. Real preferred stimuli, however, are not static but dynamic.

It is well known that STDP is sensitive to cross-correlations in its inputs at its short timescale (10–30 ms) (Kempter et al. 1999; Song et al. 2000; van Rossum et al. 2000). More specifically, correlations induce transient increases of the postsynaptic firing probability, which results in dominating LTP for the concerned inputs, while the remainders tend to be depressed. As a result, the postsynaptic neuron progressively becomes selective to the co-activation of the correlated inputs. The mechanism was found robust to spiking unreliability. Specifically, Fano Factor up to ~1 can be handled as long as the number of input is sufficient (~100), because missing and extra spikes tend to compensate (Gilson et al. 2011). All these conditions are met here (Table 2), the reason why STDP is indeed able to learn visual patterns that are consistently present in the input. In this case V1 cells tend to connect to LGN cells with aligned RF, because these cells consistently fire almost synchronously, each time a straight edge crosses their RFs. Because the cells may be in a different initial state when the edge enters their RFs, because the edge is not necessarily uniform, and because of neuronal noise, synchronisation is not perfect: the relative times have a dispersion in the 10–20 ms range (Table 2). But this falls into the STDP time window, therefore the underlying edge patterns can be learned. Again this is partly due to the sparseness of stimulations obtained with natural videos (Fig. 3).

Note that not all the V1 neurons became orientation selective. Some also developed a roughly circular symmetric connectivity (e.g. Fig. 6 RF #22). (Einhäuser et al. 2002), who used a rate-based Hebbian rule with the same videos, also obtained such round RFs. So did (Delorme et al. 2001), who used STDP with natural static images (see below). However, among all the light patterns that strongly activate RGCs, i.e. inhomogeneous surfaces, straight edges seem to be the most common, and thus orientation selective

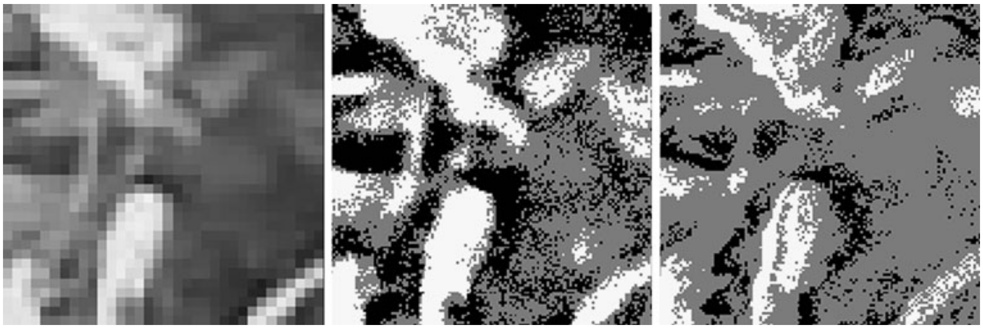

Fig. 5 One sample frame of a combined video available in online (Online Resource 1). (*Left*) Input video (*Middle*) RGC spikes emitted in a 20 ms window. White (resp. black) pixels represent ON- (resp. OFF-) cell spikes (Right) Idem for LGN spikes

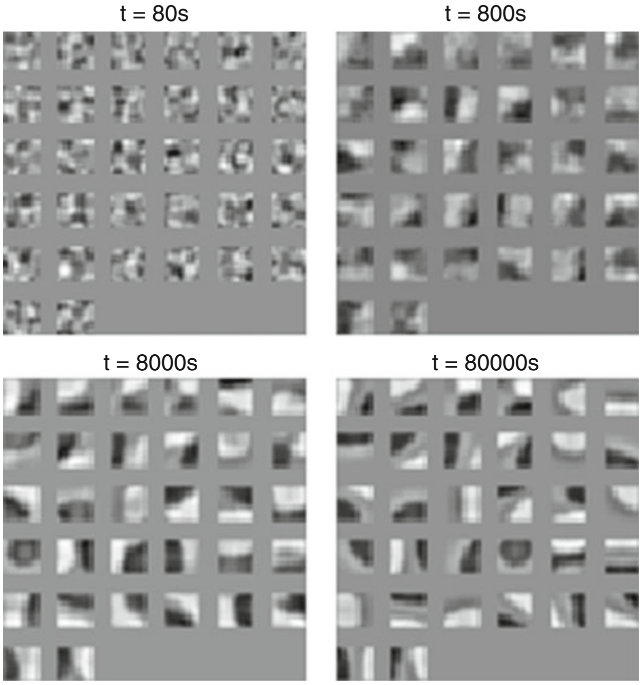

Fig. 6 Emergence of orientation selectivity. V1 RFs before training (*top left*), during training (*top right*, *bottom left*), and after convergence (*bottom right*)

RFs dominate, in line with experimentation (Singer et al. 1975).

To assess the learning robustness, we also trained V1 neurons using the fourth movie by (Betsch et al. 2004) (whereas the first one was used in the baseline simulation). We chose the fourth movie because its content (rocks, man-made buildings) is quite different from the first's (pound, trees). Yet it led to similar RFs (see Online Resource 2). We also tried to decrease the level of sparseness. Indeed, it may be argued that our baseline mean firing rates (7 Hz in the retina, and 2 Hz in the LGN, see Table 2) are weak. We thus increased the IPL gain λ_G to 300 Hz (against 150 Hz for the baseline simulation, see Table 1). That led to mean firing rates in the retina and LGN of respectively 13 Hz and 5 Hz. Not surprisingly, the crosscorrelation timescales increased in the retina: 37 ms (because the RGCs spent more time in supratheshold mode), and consequently in the LGN: 17 ms (against respectively 33 ms and 14 ms for the baseline, see Table 2). This slowed down learning, yet most of the V1 neurons still became orientation selective (see their RFs in Online Resource 3). With a much higher λ_G, learning would completely collapse. In any case, it is worth mentioning that real mean firing rates, averaged across both cells and time, under naturalistic conditions, are largely unknown, and often overestimated, because most experimentalists only use stimuli that elicit strong responses, and only record from the most responsive cells.

It had been shown before that orientation selectivity may emerge thanks to STDP, but in “discrete processing mode”, with image by image propagation and intensity-to-latency conversion (Delorme et al. 2001). The authors argued that sequential presentation of images in their model could result from saccadic eye movements. However, as mentioned in the introduction, and demonstrated in the next section, the “intensity to post-saccade latency” conversion is questionable. Furthermore, it is unclear to what extent saccades are necessary for normal RF development, and many animals make few of them, including cats (Betsch et al. 2004). We thus think that the present study—which, to our knowledge, is the first plausibility proof for STDP-based emergence of orientation selectivity with continuous spike trains produced by natural videos—is of theoretical importance. In particular, we predict that saccades are not required for RF development.

We obtained the same qualitative results with a recent similar model, yet much simplified and hardware-based (Zamarreño-Ramos et al. 2011). Specifically, we used a hardware artificial retina developed at INI Zurich (Lichtsteiner et al. 2007), which sensed the external world in a continuous (frame-free) manner, and generated spikes that were asynchronously propagated, as they flowed in, until a layer mimicking the primary visual cortex (the LGN was ignored). In this layer, neurons were equipped with memristor-based STDP, and did become orientation selective.

3.4 Absolute vs. relative spike time coding in saccadic vision

We now want to compare relative spike time coding with absolute spike time coding, which requires a reference time. As explained in the introduction, a stimulus onset may provide this reference time, but this is a rather unnatural situation. Instead, we consider the times-to-first-spike with respect to saccade landing times. More specifically, we picked one movie frame (see Fig. 7) and generated 100 saccadic image sequences from random pre-saccadic locations towards the same target zone, which contains a vertical edge. Before each saccade, the retina “looked” at the pre-saccadic location for 675 ms, so that a stationary regime was reached. This was followed by a 50 ms long saccade with constant angular velocity (~500°/s, depending on the random pre-saccadic location). The saccade landing time defines $t=0$, and we looked at the transient post-saccade activity. On half of the trials we lowered the RMS contrast to 50% of its original value (see Section 2.2). We refer to those as “low contrast trials”, while “high contrast trials” designates trials with unchanged contrast.

We first consider two adjacent vertically aligned RGC OFF cells which, given the target zone (see Fig. 7 inset), receive a similar asymptotic activation level (that is, once

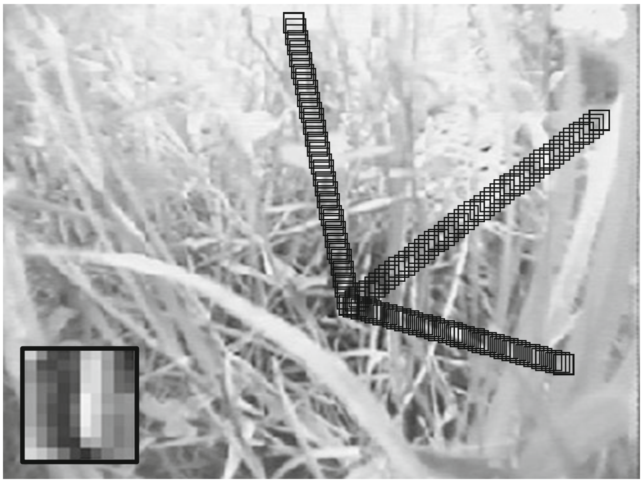

Fig. 7 Saccade illustration. Saccades start from random locations, but all land on the same image zone, which contains a vertical edge (the inset at the bottom left zooms on the landing zone). All saccades last 50 ms, and angular velocity is constant during those 50 ms (~500°/s, depending on starting location). To materialize the trajectory, one frame is drown every ms

the post-saccadic steady regime is reached). They are connected to 2 LGN OFF-cells, which thus also receive similar asymptotic activation levels.

Is time-to-first-spike coding at work in our retina model? Under the assumption that these times encode the target image and only the target image, through intensity-to-latency conversion, they should be identical for all the trials, at least within a contrast condition. Instead, as can be seen in Fig. 8a, and quantified in Table 3 (first two lines), these times are highly variable (coefficient of variation ~0.6).

Is relative spike time coding at work in our retina model? Under this hypothesis our two RGCs should fire their first spikes simultaneously (while the time they do so is irrelevant). Importantly, the above-mentioned time-to-first-spike variability has two sources: the neuronal noise (see Eq. (1)), and the different saccadic trajectories across trials. If the noise is independent between neurons, the trajectory effect is not, because the RGCs' RFs significantly overlap. As a result, the latency difference between the two cells is slightly less variable (Table 3, third line) than expected for independent random variables. Another way to see that is to compute the Pearson correlation coefficients, which are weak, but significant (Table 3, fourth line). Low contrast led to slightly longer latencies but this effect was not significant (*t*-test, $p=0.27$). All together, these results show that absolute times are highly variable, and thus poor encoders of the target image, while relative times do a slightly better job, in line with experimentation in salamander (Gollisch and Meister 2008).

Things get better in the LGN, which again increases the spike reliability and temporal precision (see Fig. 8b). It is now clear that the two cells tend to fire their first spikes at the same time, even though this time is highly variable across trials. To quantify this, we reported first spike latencies in Table 3 (these latencies do not include conduction delays RGC->LGN, and represent integration times only). The absolute latency coefficient of variation is now of about 0.4 only (lines 5 and 6). More importantly, the latency difference between the two cells is much less variable than expected for independent processes (line 7), or, equivalently, these latencies are highly correlated (last line). So once again, relative spike time coding clearly outperforms absolute spike time coding. Besides, low contrast now significantly increases latencies by ~10 ms (*t*-test, $p=10^{-7}$).

We also verified that latency differences between the two cells never reached significance, neither in the retina nor in the LGN (paired *t*-test), confirming that both cells receive very similar post-saccadic activation levels.

Finally we looked at the responses of one V1 cell (in isolation) that became selective to vertical edges with dark regions on the left and bright regions on the right (the last cell in Fig. 6). This cell thus strongly responds to the saccade landing zone (Fig. 7, inset). In addition to varying the contrast as before, we also rotated the whole stimulus sequences, in order to test the cell's orientation selectivity. We used 50 trials for each contrast x orientation condition. The corresponding raster plots are shown in Fig. 9a and b (again the latencies we report do not include conduction delays). The phenomenal speed of object recognition implies that orientations should be extracted as soon as V1 cells have fired one or two spikes, which suggests that information may be encoded in the first spike latencies (Thorpe and Imbert 1989). As can be seen in Fig. 9a and b, these latencies tend to increase (1) as contrast decreases and (2) as stimulus orientation gets further apart from the cell's preferred orientation. Both phenomena are observed experimentally: (1) in monkeys and cats (Albrecht et al. 2002; Gawne et al. 1996), and (2) in monkeys (Celebrini et al. 1993).

Figure 9c allows quantifying those latency shifts for our V1 cell. A 50% contrast decrease significantly increases the latencies by 10–20 ms, in agreement with experimentation in cats (Albrecht et al. 2002)(on this plot we chose not to represent conditions with which more than 20% of the trials elicited no spike at all, that is some extreme orientations). In addition, a shift by 45° from the cell's preferred orientation significantly increases latencies by ~5 ms at high contrast and by ~10 ms at low contrast. It is thus unlikely that the brain makes use of the absolute latencies to estimate orientations, because a non-matching orientation or a weak contrast would be undistinguishable. Instead, this suggests the following population coding scheme: when a stimulus is presented to a pool of cells with the same RF but different preferred orientations, then the stimulus orientation is given by the preferred orientation of the cells

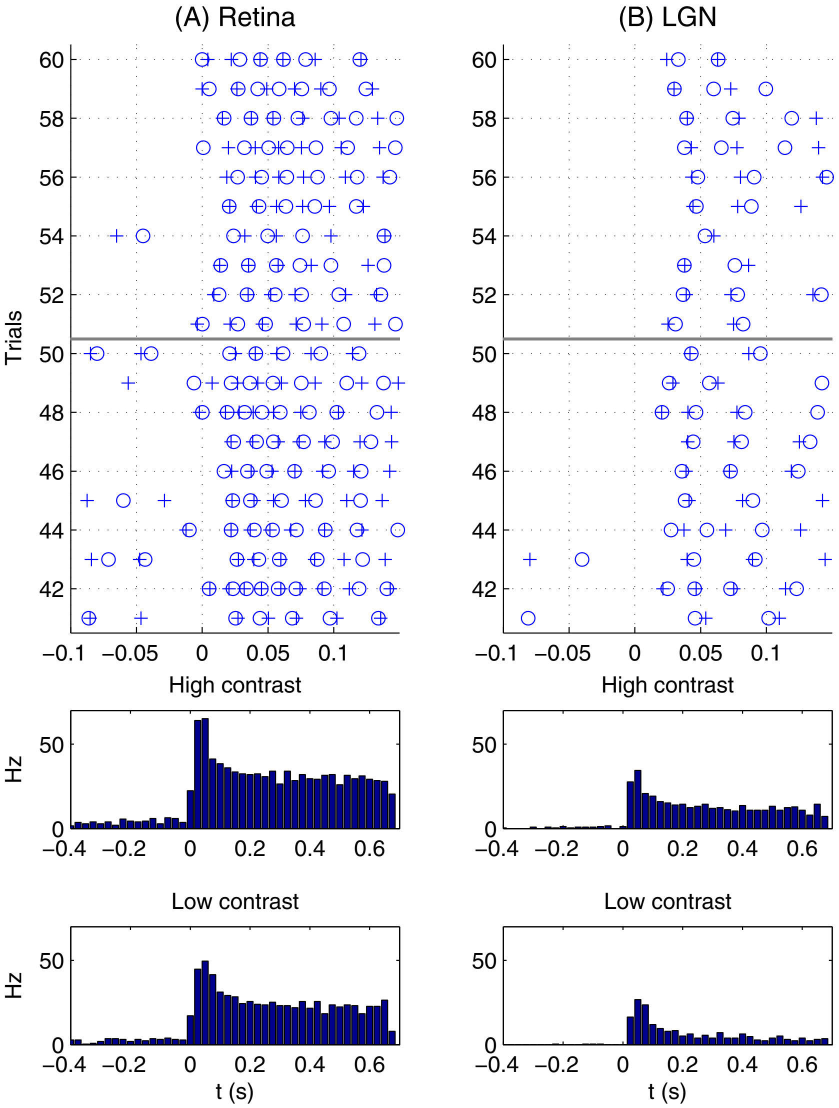

Fig. 8 Transient activity after saccade landing, in the retina (**a**) and the LGN (**b**). (Top) Raster plots with 10 high contrast trials (number≤50), and 10 low contrast trials (number>50; the horizontal grey lines separate the two contrast conditions), for two adjacent OFF-cells, with similar asymptotic activation. Cell #1 (resp. #2)'s spikes are represented with '+' (resp. 'o') (*Bottom*) Mean PSTH for the two cells with high and low contrast

that fire first. So orientation information would be encoded in the relative latencies, or in the recruiting rank ("rank order coding" (Thorpe and Gautrais 1998)). These relative latencies (or ranks) could be exploited by downstream neurons in extrastriate areas to detect more complex visual features (Masquelier and Thorpe 2007).

Table 3 Latencies (i.e. time-to-first-spike w.r.t to saccade landing time) in the retina and the LGN, for two adjacent OFF-cells (mean±standard deviation) that receive very similar post-saccadic activation levels

	Low contrast trials	High contrast trials	All trials
RGC—Latency 1	18±11 ms	17±10 ms	17±10 ms
RGC—Latency 2	19±12 ms	16±9 ms	18±11 ms
RGC—Latency difference	0±13 ms (v.s. 0±16 ms if indep.)	0±7 ms (v.s. 0±13 ms if indep.)	0±10 ms (v.s. 0±15 ms if indep.)
RGC—Latency Pearson correlation coefficient	0.32 (p=0.02)	0.76 (p=2.10^{-10})	0.5 (p=10^{-7})
LGN—Latency 1	41±13 ms	32±13 ms	37±14 ms
LGN—Latency 2	42±14 ms	31±13 ms	36±15 ms
LGN—Latency difference	0±8 ms (v.s. 0±19 ms if indep.)	1±4 ms (v.s. 0±13 ms if indep.)	0±10 ms (v.s. 0±21 ms if indep.)
LGN—Latency Pearson correlation coefficient	0.84 (p=4.10^{-14})	0.95 (p=6.10^{-25})	0.90 (p=10^{-37})

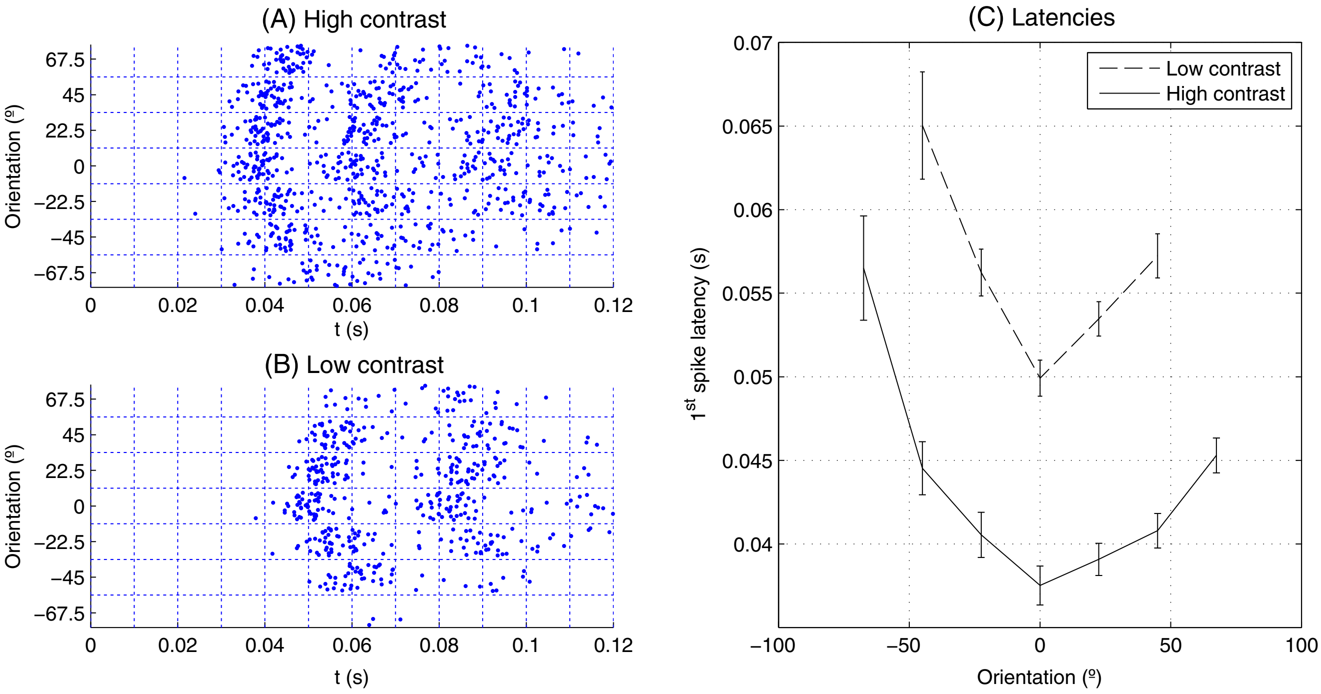

Fig. 9 Transient activity after saccade landing for a V1 cell selective to vertical edges (the last cell in Fig. 6), varying contrast and stimulus orientation. (**a**) Raster plots in the high contrast condition. There are 7 blocks of 50 trials with identical orientation (**b**) Idem in the low contrast condition (**c**) Time-to-first spike (w.r.t saccade landing) as a function of stimulus orientation, in the two contrast conditions. Error bars show 95% confidence intervals

The contrast, on the other hand, could be estimated by the mean absolute firing times. It is unclear, however, if conscious perception has access to it: it is well known that people perform poorly at absolute estimations (such as global luminance or contrast), while they are much better at discriminating stimuli (which one is the brighter/more contrasted) (see for e.g. (Stewart et al. 2005) and references there in). This suggests that for perception in general, relative latencies seem to matter more than absolute ones.

To conclude, even when reference times are available (here saccade landing times), relative spike time coding seems to outperform absolute spike time coding. And of course when no reference time is available, relative spike time coding is the only option.

4 Discussion

Our claims are four-fold. First, we predict that neighbouring RGCs in natural continuous vision have spike-time correlations at a short timescale (~30 ms), despite shifted RF, independent noise, and possible jitter accumulation. This is due to the natural world's spatio-temporal statistics, which lead to sparse responses. Second, we suggest one important role of the LGN is to narrow this crosscorrelation timescale down to ~15 ms, by filtering out some noise. Third, the output of the LGN has the reliability and temporal precision required so that downstream neurons in V1 layer 4, if equipped with STDP, can gradually become orientation selective, even in the absence of a reference time such as a saccade or stimulus onset. Fourth, if introducing saccades, whose landing times provide reference times, then relative spike times are more precise than absolute ones in both the retina and the LGN, and encode orientations more robustly in V1.

We feel that not enough attention has been paid to these relative times in the literature. Experimentalists often report *absolute* spike times, w.r.t a stimulus onset, or the peak of a local field potential oscillation for phase-of-firing coding (Montemurro et al. 2008; Masquelier et al. 2009b), but not the relative ones—which requires multiple neurons to be recorded simultaneously. Techniques to do so are now widely available (Stevenson and Kording 2011), and early studies do indicate that relative latencies are often more informative than the absolute ones in the salamander retina (Gollisch and Meister 2008), and in other modalities (Johansson and Birznieks 2004; Chase and Young 2007; Panzeri and Diamond 2010). This may be because neuronal noise is often correlated across neurons. Consequently, it may affect absolute latencies similarly, and thus have a weak impact on relative ones.

Our model suggests that neither the retina nor the LGN de-correlate much their inputs, in accordance with experimentation (Puchalla et al. 2005; Desbordes et al. 2008).

Instead, in our model, de-correlation takes place later, in V1: it is well known that edge filters similar to those of Fig. 6 are the independent components of natural images (Bell and Sejnowski 1997; van Hateren and Ruderman 1998; van Hateren and van der Schaaf 1998). This is consistent with experiments in macaque V1 showing that responses to natural images are de-correlated (Vinje and Gallant 2000), although the authors did not record multiple cells simultaneously. V1 neurons thus reduce the redundancy present in the natural environment in order to build a compact, economic, representation (Barlow 1961). Why is not this done earlier, in the retina or the LGN? We speculate that redundancy could be kept high until V1 to handle noisy signal propagation from the eyes to the cortex.

Classic rate-based hebbian learning can also lead to V1-like RFs (e.g. (Miller and MacKay 1994; Einhäuser et al. 2002)), also by learning the input correlation structure. Those models, however: (1) capture only mean firing rates, not individual spikes, and hence are agnostic about spike time related questions, such as the plausibility of time-to-first spike coding or rank order coding, or of STDP as mechanism to account for RF development; and: (2) are steady, and thus may not capture some key aspects of visual processing (we will come back to this point). Our neurophysiologically-plausible approach also contrasts with objective function approaches, which show that V1-like RFs optimize certain functions (such as sparseness (Olshausen and Field 1996; Rehn and Sommer 2007), statistical independence (Bell and Sejnowski 1997; van Hateren and Ruderman 1998; van Hateren and van der Schaaf 1998; Hyvärinen and Hoyer 2001) or temporal continuity and slowness (Wiskott and Sejnowski 2002; Körding et al. 2004; Berkes and Wiskott 2005), but do not address the issue of how the RFs could develop.

One important limitation of our work, however, is that we ignored feedback, following Hubel and Wiesel's model of feedforward orientation selectivity in the early visual system (Hubel and Wiesel 1962). While there is no feedback in the retina, relay neurons in the LGN do receive, in addition to direct visual input predominantly from a single RGC, indirect input from other sources including interneurons, thalamic reticular nucleus, and the visual cortex. This indirect output, however, is only responsible for 5% of the response variance on average, and for a maximum of 25% in some cells (Babadi et al. 2010). Furthermore, one of the main effects of cortico-thalamic projection may be to increase spike time precision in the LGN (Wörgötter et al. 1998). Thus if including such projections, we speculate that the mechanisms that we propose here for orientation learning and decoding in V1 would only work better.

Another limitation is that unlike real RGCs, ours are homogeneous in terms of spatial and temporal processing scales. Cross-correlations would decrease if using multiple scales. In this case, we speculate that STDP would tend to pool neurons with similar scales, because those would be more correlated. Future work will test this hypothesis.

Besides our four main claims, we want to stress that vision is a continuous and dynamic process; therefore models should be continuous and dynamic as well. Steady models do not capture motions and transients, which are nevertheless crucial for perception, even of still images, as mentioned in the introduction. Besides, only continuous dynamic models can simulate the experimental protocols of Rapid Serial Visual Presentation (RSVP) and visual masking. To account for high level object and scene recognition these models should include many more layers than in this early study, and simulate the extrastriate visual cortex. The timescales involved in natural continuous vision processing are fast (~10 ms) (Butts et al. 2007), and individual neurons' firing rates are not well defined at such a fine temporal resolution, so spiking neuron models should be preferred. STDP, which is able to detect consistently repeating spike patterns even in continuous activity (Masquelier et al. 2008), probably plays a key role at all stages.

Acknowledgements This research was supported by the Fyssen Foundation and the FP7 European Project Coronet. We thank Adrien Wohrer for having developed the Virtual Retina simulator (Wohrer and Kornprobst 2009), both user-friendly and highly configurable, and for the quality of his support. We also thank Wolfgang Einhäuser for kindly providing the videos used in (Betsch et al. 2004), and Mario Pannunzi for the numerous insightful brainstorms we had.

References

Albrecht, D. G., Geisler, W. S., Frazor, R. A., & Crane, A. M. (2002). Visual cortex neurons of monkeys and cats: temporal dynamics of the contrast response function. *Journal of Neurophysiology, 88*(2), 888–913.

Babadi, B., Casti, A., Xiao, Y., Kaplan, E., & Paninski, L. (2010). A generalized linear model of the impact of direct and indirect inputs to the lateral geniculate nucleus. *Journal of Vision, 10*(10), 22.

Bacon-Mace, N., Mace, M. J., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research, 45*(11), 1459–69.

Barlow, H. (1961). Possible principles underlying the transformation of sensory messages. In *Sensory communication* (pp. 217–234). Cambridge: MIT. wa rosenblith edition.

Bell, A. J., & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research, 37*(23), 3327–3338.

Berkes, P., & Wiskott, L. (2005). Slow feature analysis yields a rich repertoire of complex cell properties. *Journal of Vision, 5*(6), 579–602.

Betsch, B., Einhäuser, W., Körding, K., & König, P. (2004). The world from a cat's perspective—statistics of natural videos. *Biological Cybernetics, 90*(1), 41–50.

Brette, R., & Guigon, E. (2003). Reliability of spike timing is a general property of spiking model neurons. *Neural Computation, 15*(2), 279–308.

Butts, D. A., Weng, C., Jin, J., Yeh, C.-I., Lesica, N. A., Alonso, J.-M., et al. (2007). Temporal precision in the neural code and the timescales of natural vision. *Nature, 449*(7158), 92–95.

Cai, D., DeAngelis, G. C., & Freeman, R. D. (1997). Spatiotemporal receptive field organization in the lateral geniculate nucleus of cats and kittens. *Journal of Neurophysiology, 78*(2), 1045–1061.

Caporale, N., & Dan, Y. (2008). Spike timing-dependent plasticity: a hebbian learning rule. *Annual Review of Neuroscience, 31*, 25–46.

Carandini, M., Horton, J. C., & Sincich, L. C. (2007). Thalamic filtering of retinal spike trains by postsynaptic summation. *Journal of Vision, 7*(14), 20.1–2011.

Celebrini, S., Thorpe, S., Trotter, Y., & Imbert, M. (1993). Dynamics of orientation coding in area V1 of the awake primate. *Visual Neuroscience, 10*(5), 811–825.

Chapman, B., Zahs, K. R., & Stryker, M. P. (1991). Relation of cortical cell orientation selectivity to alignment of receptive fields of the geniculocortical afferents that arborize within a single orientation column in ferret visual cortex. *Journal of Neuroscience, 11*(5), 1347–1358.

Chase, S. M., & Young, E. D. (2007). First-spike latency information in single neurons increases when referenced to population onset. *Proceedings of the National Academy of Sciences of the United States of America, 104*(12), 5175–5180.

Chung, S., & Ferster, D. (1998). Strength and orientation tuning of the thalamic input to simple cells revealed by electrically evoked cortical suppression. *Neuron, 20*(6), 1177–1189.

Coppola, D., & Purves, D. (1996). The extraordinarily rapid disappearance of entopic images. *Proceedings of the National Academy of Sciences of the United States of America, 93*(15), 8001–8004.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: face detection in just 100 ms. *Journal of Vision, 10* (4), 1–17.

Delorme, A., Perrinet, L., Thorpe, S., & Samuelides, M. (2001). Networks of integrate-and-fire neurons using rank order coding B: spike timing dependent plasticity and emergence of orientation selectivity. *Neurocomputing, 38–40*, 539–545.

Delorme, A., & Thorpe, S. J. (2001). Face identification using one spike per neuron: resistance to image degradations. *Neural Networks, 14*(6–7), 795–803.

Desbordes, G., Jin, J., Weng, C., Lesica, N. A., Stanley, G. B., & Alonso, J.-M. (2008). Timing precision in population coding of natural scenes in the early visual system. *PLoS Biology, 6*(12), e324.

Einhäuser, W., Kayser, C., König, P., & Körding, K. P. (2002). Learning the invariance properties of complex cells from their responses to natural stimuli. *European Journal of Neuroscience, 15*(3), 475–486.

Enroth-Cugell, C., Robson, J. G., Schweitzer-Tong, D. E., & Watson, A. B. (1983). Spatio-temporal interactions in cat retinal ganglion cells showing linear spatial summation. *The Journal of Physiology, 341*, 279–307.

Fabre-Thorpe, M., Richard, G., & Thorpe, S. J. (1998). Rapid categorization of natural images by rhesus monkeys. *NeuroReport, 9*(2), 303–8.

Ferster, D., Chung, S., & Wheat, H. (1996). Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature, 380* (6571), 249–252.

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation, 3*, 194–200.

Fukushima, K. (1980). Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics, 36*(4), 193–202.

Gawne, T., Kjaer, T., & Richmond, B. (1996). Latency: another potential code for feature binding in striate cortex. *Journal of Neurophysiology, 76*(2), 1356–1360.

Gerstner, W., Ritz, R., & van Hemmen, J. L. (1993). Why spikes? hebbian learning and retrieval of time-resolved excitation patterns. *Biological Cybernetics, 69*(5–6), 503–515.

Gilson, M., Masquelier, T., & Hugues, E. (2011). STDP allows fast rate-modulated coding with Poisson-like spike trains. *PLoS Computational Biology (in press)*.

Girard, P., Jouffrais, C., & Kirchner, C. H. (2008). Ultra-rapid categorisation in non-human primates. *Animal Cognition, 11*(3), 485–493.

Gollisch, T., & Meister, M. (2008). Rapid neural coding in the retina with relative spike latencies. *Science, 319*(5866), 1108–1111.

Guyonneau, R., VanRullen, R., & Thorpe, S. (2005). Neurons tune to the earliest spikes through STDP. *Neural Computation, 17*(4), 859–879.

Hubel, D., & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106–154.

Hung, C., Kreiman, G., Poggio, T., & DiCarlo, J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310*(5749), 863–866.

Hyvärinen, A., & Hoyer, P. O. (2001). A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Research, 41*(18), 2413–2423.

Johansson, R. S., & Birznieks, I. (2004). First spikes in ensembles of human tactile afferents code complex spatial fingertip events. *Nature Neuroscience, 7*(2), 170–177.

Kara, P., Reinagel, P., & Reid, R. C. (2000). Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron, 27*(3), 635–646.

Keat, J., Reinagel, P., Reid, R. C., & Meister, M. (2001). Predicting every spike: a model for the responses of visual neurons. *Neuron, 30*(3), 803–817.

Kempter, R., Gerstner, W., & van Hemmen, J. L. (1999). Hebbian learning and spiking neurons. *Physical Review E, 59*(4), 4498–4514.

Kirchner, H., & Thorpe, S. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Research, 46*(11), 1762–1776.

König, P., Engel, A. K., & Singer, W. (1996). Integrator or coincidence detector? The role of the cortical neuron revisited. *Trends in Neurosciences, 19*(4), 130–7.

Körding, K., Kayser, C., Einhäuser, W., & König, P. (2004). How are complex cell properties adapted to the statistics of natural stimuli? *Journal of Neurophysiology, 91*(1), 206–212.

LeCun, Y., & Bengio, Y. (1998). Convolutional networks for images, speech, and time series. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 255–258). Cambridge: MIT.

Lichtsteiner, P., Posch, C., & Delbruck, T. (2007). An 128×128 120db 15us-latency temporal contrast vision sensor. *IEEE Journal of Solid-State Circuits, 43*(2), 566–576.

Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron, 62* (2), 281–290.

Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience, 5*(3), 229–240.

Masquelier, T., Guyonneau, R., & Thorpe, S. J. (2008). Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains. *PloS One, 3*(1), e1377.

Masquelier, T., Guyonneau, R., & Thorpe, S. J. (2009). Competitive STDP-based spike pattern learning. *Neural Computation, 21*(5), 1259–1276.

Masquelier, T., Hugues, E., Deco, G., & Thorpe, S. J. (2009). Oscillations, phase-of-firing coding, and spike timing-dependent plasticity: an efficient learning scheme. *Journal of Neuroscience, 29*(43), 13484–13493.

Masquelier, T., Serre, T., Thorpe, S., & Poggio, T. (2007). Learning complex cell invariance from natural videos: a plausibility proof. *Massachusetts Institute of Technology*, CBCL Paper #269/MIT-CSAIL-TR #2007-060.

Masquelier, T., & Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Computational Biology, 3*(2), e31.

Miller, K. D., & MacKay, D. J. C. (1994). The role of constraints in hebbian learning. *Neural Computation, 6*, 100–126.

Montemurro, M. A., Rasch, M. J., Murayama, Y., Logothetis, N. K., & Panzeri, S. (2008). Phase-of-firing coding of natural visual stimuli in primary visual cortex. *Current Biology, 18*(5), 375–380.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381*, 607–609.

Oram, M., & Perrett, D. (1992). Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology, 68*(1), 70–84.

Panzeri, S., & Diamond, M. E. (2010). Information carried by population spike times in the whisker sensory cortex can be decoded without knowledge of stimulus time. *Frontiers in Synaptic Neuroscience, 2*(17), 1–14.

Puchalla, J. L., Schneidman, E., Harris, R. A., & Berry, M. J. (2005). Redundancy in the population code of the retina. *Neuron, 46*(3), 493–504.

Rathbun, D. L., Warland, D. K., & Usrey, W. M. (2010). Spike timing and information transmission at retinogeniculate synapses. *Journal of Neuroscience, 30*(41), 13558–13566.

Rehn, M., & Sommer, F. T. (2007). A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. *Journal of Computational Neuroscience, 22*(2), 135–146.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019–1025.

Rolls, E., & Milward, T. (2000). A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural Computation, 12*(11), 2547–2572.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5*(7), 629–30.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proc. Nat. Acad. Sci. USA*, 104 (15).

Singer, W., Tretter, F., & Cynader, M. (1975). Organization of cat striate cortex: a correlation of receptive-field properties with afferent and efferent connections. *Journal of Neurophysiology, 38* (5), 1080–1098.

Song, S., Miller, K., & Abbott, L. (2000). Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience, 3*(9), 919–926.

Spratling, M. (2005). Learning viewpoint invariant perceptual representations from cluttered images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(5).

Stevenson, I. H., & Kording, K. P. (2011). How advances in neural recording affect data analysis. *Nature Neuroscience, 14*(2), 139–142.

Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review, 112* (4), 881–911.

Stone, J. (1965). A quantitative analysis of the distribution of ganglion cells in the cat's retina. *The Journal of Comparative Neurology, 124*(3), 337–352.

Stringer, S., & Rolls, E. (2000). Position invariant recognition in the visual system with cluttered environments. *Neural Networks, 13* (3), 305–315.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–2.

Thorpe, S., & Gautrais, J. (1998). Rank order coding. In J. M. Bower (Ed.), *Computational neuroscience: Trends in research* (pp. 113–118). New York: Plenum.

Thorpe, S., & Imbert, M. (1989). Biological constraints on connectionist modelling. In R. Pfeifer, Z. Schreter, F. Fogelman-Soulie, & L. Steels (Eds.), *Connectionism in perspective* (pp. 63–92). Amsterdam: Elsevier.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5*(7), 682–687.

van Hateren, J. H., & Ruderman, D. L. (1998). Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proceedings. Biological sciences / The Royal Society, 265*(1412), 2315–2320.

van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings. Biological sciences / The Royal Society, 265*(1394), 359–366.

van Rossum, M. C., Bi, G. Q., & Turrigiano, G. G. (2000). Stable hebbian learning from spike timing-dependent plasticity. *Journal of Neuroscience, 20*(23), 8812–8821.

VanRullen, R., Gautrais, J., Delorme, A., & Thorpe, S. (1998). Face processing using one spike per neurone. *Biosystems, 48*(1–3), 229–239.

VanRullen, R., & Thorpe, S. (2001). Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Computation, 13*(6), 1255–1283.

VanRullen, R., & Thorpe, S. (2002). Surfing a spike wave down the ventral stream. *Vision Research, 42*(23), 2593–2615.

Vinje, W. E., & Gallant, J. L. (2000). Sparse coding and decorrelation in primary visual cortex during natural vision. *Science, 287* (5456), 1273–1276.

Wallis, G., & Rolls, E. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology, 51*(2), 167–194.

Williams, P. E., Mechler, F., Gordon, J., Shapley, R., & Hawken, M. J. (2004). Entrainment to video displays in primary visual cortex of macaque and humans. *Journal of Neuroscience, 24*(38), 8278–8288.

Wilson, J. R., & Sherman, S. M. (1976). Receptive-field characteristics of neurons in cat striate cortex: Changes with visual field eccentricity. *Journal of Neurophysiology, 39*(3), 512–533.

Wiskott, L., & Sejnowski, T. J. (2002). Slow feature analysis: unsupervised learning of invariances. *Neural Computation, 14* (4), 715–770.

Wohrer, A. (2008). *Model and large-scale simulator of a biological retina, with contrast gain control*. PhD thesis, University of Nice-Sophia Antipolis.

Wohrer, A., & Kornprobst, P. (2009). Virtual retina: a biological retina model and simulator, with contrast gain control. *Journal of Computational Neuroscience, 26*(2), 219–249.

Wörgötter, F., Nelle, E., Li, B., & Funke, K. (1998). The influence of corticofugal feedback on the temporal structure of visual responses of cat thalamic relay cells. *The Journal of Physiology, 509*(Pt 3), 797–815.

Zamarreño-Ramos, C., Camuñas-Mesa, L., Perez-Carrasco, J. A., Masquelier, T., Serrano-Gotarredona, T., & Linares-Barranco, B. (2011). On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex. *Front. Neurosc.—Neuromorph. Eng.*, 5(26).