Learning to recognize objects using waves of spikes and Spike Timing-Dependent Plasticity

Timothée Masquelier, Simon J Thorpe

Learning to recognize objects using waves of spikes and Spike Timing-Dependent Plasticity

Timothée Masquelier and Simon J. Thorpe

Abstract— **This paper focuses on feedforward spiking neuron models of the visual cortex. Essentially, we show that a combination of a temporal coding scheme where the most strongly activated neurons fire first with Spike Timing-Dependent Plasticity leads to a situation where neurons will gradually become selective to visual patterns that are both salient, and consistently present in the inputs. At the same time, their responses become more and more rapid. These responses can then be used very effectively to perform object recognition in natural images.**

We firmly believe that such mechanisms are a key to understanding the remarkable efficiency of the primate visual system, and that similar mechanisms could and should be implemented in artificial vision systems, possibly using Address Event Representation (AER) and memristors.

Subsequent work will explore video processing, the use of feedback connections, and oscillatory regimes.

I. Introduction

A. Spike waves in the visual system

Speed of recognition in the primate visual system imposes severe constraints on the underlying neuronal processes. There is now considerable behavioral [1]–[7] and electrophysiological [8]–[11] evidence that the primate visual system can achieve high level object recognition in 80-100ms. Given that about 10 neuronal layers are involved in that kind of processing, the time-window available for each neuron to perform computation is only about 10ms, and, given that the firing rates are barely above 100Hz in the visual system, such a small window will typically contain at most one spike [12]. A classical rate coding scheme, in which individual neurons encode information in their mean firing rate, is thus effectively ruled out. Instead, the information could be encoded in which afferents were recruited, and possibly additionally in the relative recruiting times, a scheme referred to as 'rank order coding' [13]. Note that if computation is restricted to one spike per neuron, the use of feedback loops is also effectively ruled out. This means the first spike wave after stimulus onset probably does much more than conventionally assumed, and it is this sort of rapid processing that we are interested in this paper.

Specifically, we propose to model the visual system as a feedforward spiking neural network that operates in the temporal domain. Images are presented one by one to the network. The first layer performs convolutions on each of them, and applies an intensity-to-latency conversion on the result (see Fig. 1): the more a neuron is stimulated (for example by the presence of a salient edge in its receptive field), the earlier it fires a spike. This intensity-to-latency conversion is in accordance with recordings in the primary visual cortex (V1) showing that response latency decreases with stimulus contrast [14], [15] and with the proximity between the stimulus orientation and the neuron's preferred orientation [16]. These spikes are then propagated asynchronously through the feedforward network. We restrict the computation to one spike per neuron, which leads to efficient implementations.

Several nice properties come for free with this 'time-to-first-spike' coding scheme. First, Winner-Take-All (WTA) mechanisms are easy to implement: with time-to-first-spike coding kWTA simply means that the earliest firing neurons should prevent their competitors from firing for a while - something that lateral inhibition can easily do [17]. Such a mechanism can be efficiently implemented in hardware, as we will see below. Neural circuits to implement WTA in rate-based coding frameworks (and the particular case of 1WTA, *i.e.* max operation) have also been proposed [18]–[24], but these circuits are significantly more complex, and their responses are usually unreliable before a steady regime is reached, which typically takes several tens of milliseconds, especially if input values are close to each other.

Second, in our framework the earliest spikes correspond to the most salient regions of an image, and are thus usually the most informative. It follows a natural image compression scheme: VanRullen & Thorpe demonstrated, using Difference-of-Gaussian (DoG) filters, that reasonably good image reconstruction can be achieved by the time only 1% of the DoG units have fired, with little need to wait for later responses [25].

Third, if instead of considering the absolute latencies, one assumes that the information is contained in the recruiting ranks, then the resulting coding scheme is invariant to both image luminance and constrast [17].

B. Spike Timing-Dependent Plasticity

Spike Timing-Dependent Plasticity (STDP) provides an appealing mechanism for unsupervised learning in a spiking neural network. STDP is a physiological mechanism of activity-driven synaptic regulation, where an excitatory synapse receiving a spike before a postsynaptic one is emitted is reinforced (Long-Term Potentiation, LTP) whereas its strength is weakened the other way around (Long-Term

Timothée Masquelier is with the Theoretical and Computational Neuroscience Group, Departament de Tecnologia, Universitat Pompeu Fabra, 08018 Barcelona, Spain (phone: +34 93 542 14 52; email: timothee.masquelier@alum.mit.edu).

Simon J. Thorpe is with the Centre de Recherche Cerveau et Cognition Université Toulouse 3, Centre National de la Recherche Scientifique, Faculté de Médecine de Rangueil, 31062 Toulouse, France (phone: +33 5 62 17 28 03; email: simon.thorpe@cerco.ups-tlse.fr).

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

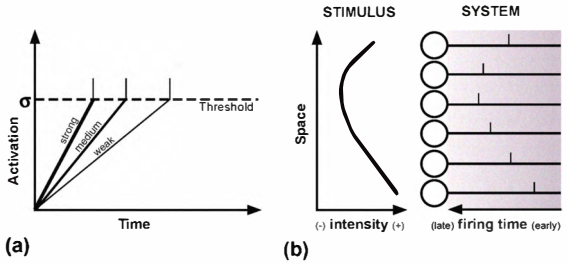

Fig. 1. Intensity-to-latency conversion. (a) For a single neuron, the weaker the stimulus, the longer the time-to-first-spike. (b) When presented to a population of neurons, the stimulus evokes a spike wave, within which asynchrony encodes the information (reproduced with permission from [26])

Depression, LTD). STDP has been observed both in vivo and in vitro, in many species (from insect to mammals) and in many brain areas (see [27] for a recent review). Note that STDP is in agreement with Hebb's postulate [28] because it reinforces the connections with the presynaptic neurons that fired slightly before the postsynaptic neuron, which are those that 'took part in firing it'.

An additive exponential update rule (see Fig. 2):

$$\Delta w_{ij} = \begin{cases} a^+ \cdot \exp\left(\frac{t_j - t_i}{\tau^+}\right) & \text{if} \quad t_j - t_i \leq 0 \quad \text{(LTP)} \\ a^- \cdot \exp\left(-\frac{t_j - t_i}{\tau^-}\right) & \text{if} \quad t_j - t_i > 0 \quad \text{(LTD)} \end{cases}$$

with the time constants $\tau^+ = 17$ ms and $\tau^- = 34$ ms, provides a reasonable approximation of the synaptic modification observed experimentally [29].

Multiplicative STDP, in which the weight updates also depend on the current weight, have also been proposed [30], [31]. Implementations also differ in the number of spike pairs they take into account: the so called 'all-to-all' mode considers all pairs of pre- and post-synaptic while the so called 'nearest spike' approximation assumes that only the nearest spikes matter (see [32] for a comparison of these two approaches).

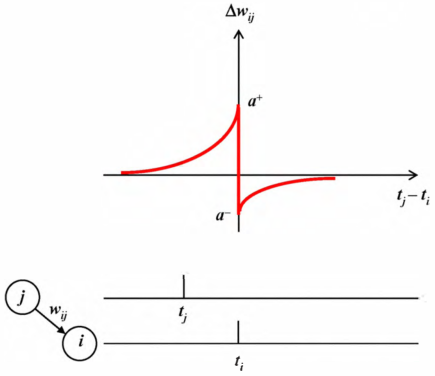

Fig. 2. Here we plotted the STDP additive synaptic weight updates as a function of the difference between the presynaptic spike time t_j and the postsynaptic spike time t_i plotted. The left part corresponds to Long Term Potentiation (LTP) and the right part to Long Term Depression (LTD).

STDP has received considerable interest from the modeling community over the last decade. In an influential paper Song *et al.* demonstrated the *competitive* nature of STDP: synapses compete for the control of the postsynaptic spikes [33]. This competition stabilizes the synaptic weights: because not all the synapses can 'win' (*i.e.* be reinforced) the sum of the synaptic weights is naturally bounded, without the need for additional normalization mechanisms. Furthermore, when the system is repeatedly presented with similar spike patterns, the winning synapses are those through which the earliest spikes arrive (on average). The ultimate effect of this synaptic modification is to make the postsynaptic neuron respond more quickly.

Gerstner & Kistler reproduced those main results and also demonstrated that STDP increased the postynaptic spike time precision by selecting inputs with low time jitter [34].

Guyonneau *et al.* tested the robustness of the results of Song *et al.* [33] in more challenging conditions, including spontaneous activity or jitter. Furthermore they also demonstrated that STDP favors inputs with short latencies, more than inputs with high firing rates or synchrony [35].

In this paper we propose to combine the two above mentioned ideas: namely time-to-first-spike coding and STDP. Section II focuses on holistic schemes, while Section III deals with part-based schemes. We then talk about hardware implementations in Section IV. Finally, we discuss related approaches and give perspectives in Section V.

II. STDP-BASED HOLISTIC LEARNING

In this section we show how STDP can lead to holistic recognition, *i.e.* recognition in only one template-matching step.

A. Single image, single neuron

Let us start with the simplest set up one can think of: an entry layer covers the whole input image and performs convolutions on it. The filters may be for example DoG (to mimic retinal ganglion cells) or Gabor-like oriented edge detectors (to mimic V1 neurons). As explained above (see Fig. 1), the convolution results are converted into spike latencies. The resulting spike train is fed into one downstream neuron equipped with STDP. Guyonneau *et al.* investigated what happens if the same image is repeatedly presented to such a system, starting from random synaptic weights [26]. Fig. 3 illustrates their results: STDP progressively concentrates synaptic weights on the earliest firing afferents, with the result that the postsynaptic spike latency decreases, until a minimum is reached. At this stage, the neuron's preferred stimulus, which can be linearly reconstructed from the synaptic weights, corresponds to the salient edges of the input images.

B. Multiple images, multiple neurons in competition

Now say you want to learn several input images. A natural extension to the above-mentioned scheme is to have a population of neurons, all integrating the image spike trains in parallel. To prevent two neurons from learning the same image, one can implement lateral inhibitory connections,

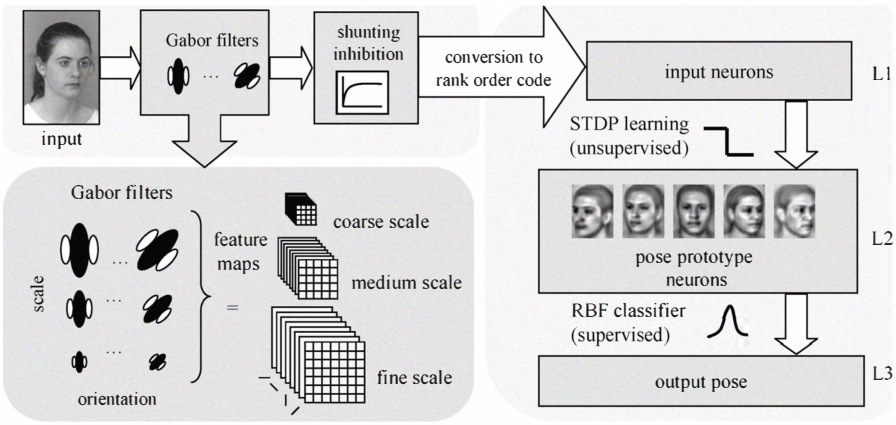

Fig. 4. Overview of the model used in [36]. The model consists of three major processing steps. Beginning with a preprocessing stage low level features are extracted and further converted into a rank-order code. Following, pose prototypes are learned from these temporal codes by applying STDP. After learning has converged, an RBF classifier is used to evaluate the responses from prototypical pose neurons (reproduced with permission from [36]).

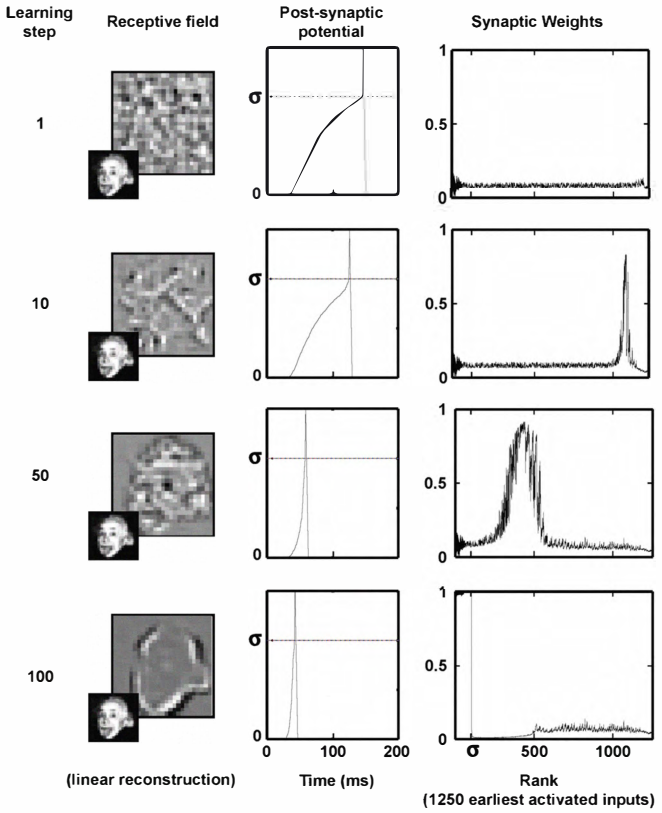

Fig. 3. One STDP neuron learns one V1-filtered image (here, Einstein's face). (Right column) Synpatic weights. Afferents are ordered with increasing latencies. Notice how STDP tracks back through the spike train, depressing synapses it had previously reinforced, until having concentrated all the weights on the earliest firing afferents. The final number of selected afferents depends on the threshold σ. (Middle column) Postsynaptic membrane potential as a function of time. Notice how the postsynaptic spike latency decreases as learning progresses (Left column) Linear reconstruction of the neuron's preferred stimulus. Before learning, since the synaptic weights are random, so is the reconstruction. But after learning, the neuron became selective to the most salient edges of the input image. Reproduced with permission from [26].

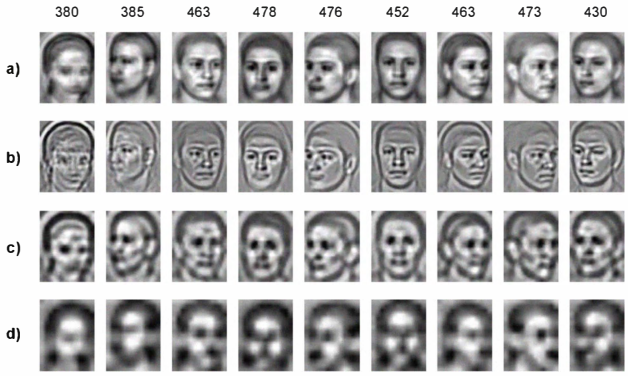

Fig. 5. Receptive fields of all STDP neurons after 4000 learning steps. Images in a) show additive reconstructions from learned weights of all feature channels (scale, orientation and phase polarity). b) - d) Idem, but separated for each Gabor scale from fine to coarse. It is clearly visible that the neurons show a preference for particular head poses. The numbers over each column denote the quantity of learning steps for each neuron (reproduced with permission from [36]).

such that as soon as one neuron fires, it strongly inhibits its neighbours, which thus cannot fire for a while (as discussed above, this is a simple and biologically plausible implementation of the well known Winner-Take-All mechanism). As a result, the neuron population then self-organizes: each neuron tends to learn a different image, or a different set of similar images. Ref. [26], [36], [37] used this approach.

For example Weidenbacher & Neumann demonstrated how such a set up can handle head pose recognition [36]. As input of the model they used images of 200 subjects in 9 different poses taken from the FERET database [38]. Those images were convolved with a set of Gabor filters at 8 orientations, 3 scales and 2 phases (ON/OFF centre receptive field, responding to positive and negative local contrast), resulting in 48 feature maps (see Fig. 4). The resulting spike trains were fed into 9 neurons in parallel, equipped with

STDP. Fig. 5 shows that after 4,000 presentations of faces in arbitrary poses, those neurons showed a clear preference for specific head poses.

It is important to note that up to this point, the learning was fully unsupervised. No external teacher signal was given to the model. Consequently, the model had no knowledge about which prototype neuron is related to which pose. Poses were only learned due to statistical regularities in the dataset.

The vector of STDP neuron responses was then fed into a Radial Basis Function (RBF) classifier, previously trained on a supervised manner, using labeled examples from the FERET database. A standard cross-validation procedure was used, and the poses of 94.7% of previously unseen examples were correctly estimated within a +/- 15° range.

III. STDP-BASED LOCAL FEATURE LEARNING

To achieve robust object recognition, despite eventual occlusions, and variations in view point, lighting conditions *etc.*, while avoiding a combinatorial explosion, it is generally useful to recognize an object as a combination of local features, as opposed to in a holistic manner. This strategy is largely employed by the brain. Specifically, the ventral stream of the visual system, involved in object recognition, is roughly hierarchically organized. At each stage neurons respond to combinations of simpler features, encoded at the preceding stage. Thus along the hierarchy neurons respond to more and more complex visual features. In an attempt to mimic this hierarchical organization a number of models have been proposed [39]–[48]. In this section we show that STDP provides an appealing mechanism for unsupervised learning in such networks.

A. Low level features

At the bottom of the hierarchy, in the primary visual cortex (V1), neurons are selective to simple oriented bars. Their selectivity can be well fitted by Gabor filters.

Delorme *et al.* showed how these Gabor-like selectivities can be obtained by applying STDP on spike trains coming from retinal ON- and OFF-center cells, modeled as DoG filters, and performing an intensity-to-latency conversion [49]. Again a 1-WTA mechanism ensured that only the first neuron to fire at each location would undergo weight modifications.

After propagating 2590 natural images they observed different sorts of selectivity that included contour orientation, end-stop and blob cells (see Fig. 6).

B. High level features

Further up in the hierarchy, say in the Infero Temporal cortex (IT), neurons respond to more complex features, such as faces or face parts. In an attempt to explain how such complex selectivities can be learned with STDP, we used a five-layer hierarchical network largely inspired by the model formerly known as HMAX [43], [48] (see Fig. 7). Specifically, we also alternated simple cells that gain selectivity through a sum operation, and complex cells that gain shift and scale invariance through a max operation.

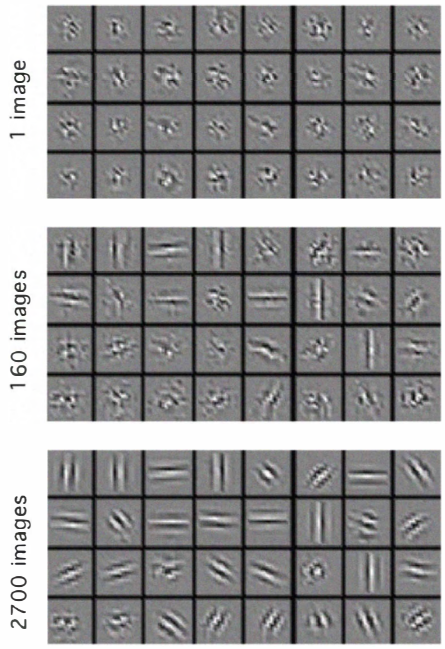

Fig. 6. Pattern of connectivity of 64 maps of V1 neurons inhibiting each other. The pattern of connectivity of neurons were initialized with two arrays (ON and OFF) of size 11x11 whose synaptic weights were distributed according to a randomized Gaussian function (the reconstruction process display here take into account both of these arrays, synapses from ON-center neurons being counted as positive and those from OFF-center neurons being counted as negative). After propagating thousands of images, coherent receptive fields were found to arise naturally for these 64 maps of neurons. Interestingly, they look like the selectivities found in V1 (reproduced with permission from [49]).

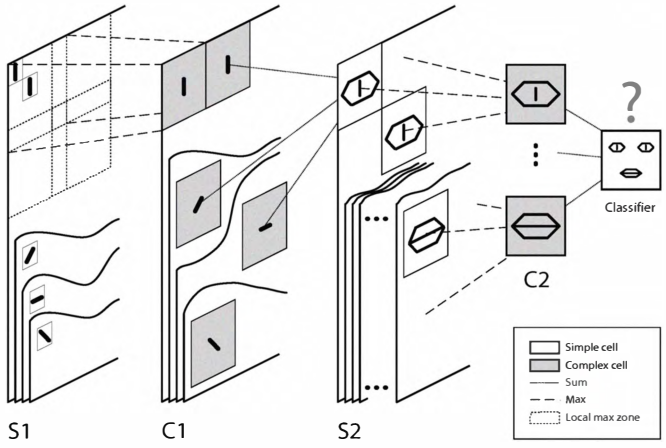

Fig. 7. Overview of our 5 layer feedforward spiking neural network. Cells are organized in retinotopic maps until the S_2 layer (included). S_1 cells detect edges. C_1 maps sub-sample S_1 maps by taking the maximum response over a square neighborhood. S_2 cells are selective to intermediate complexity visual features, defined as a combination of oriented edges (here we symbolically represented an eye detector and a mouth detector). There is one S_1–C_1–S_2 pathway for each processing scale (not represented on the figure). Then C_2 cells take the maximum response of S_2 cells over all positions and scales and are thus shift and scale invariant. Finally, a classification is done based on the C_2 cells' responses (here we symbolically represented a face / non face classifier). Here we focus on the learning of C_1 to S_2 synaptic connections through STDP. Fig. 8 shows an example of resulting selectivities with faces.

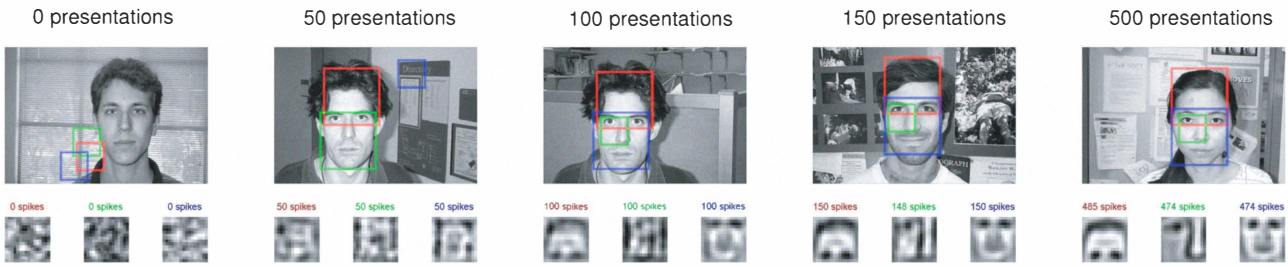

Fig. 8. Preferred stimulus reconstructions of C_2 cells after 0, 50, 150, and 500 presentations. At the top of each frame the input image is shown, with red, green or blue squares indicating the receptive fields of the cells that fired (if any). At the bottom we reconstructed the preferred stimuli of the three cells: the left cell gradually becomes selective to foreheads, the middle one to noses and left eyes, and the right one to a global view of a face. Above each reconstruction the number of postsynaptic spikes emitted is shown with the corresponding color. All the numerical parameters are the same as in [47]. See also the videos available online at: `http://dx.doi.org/10.1371/journal.pcbi.0030031`

However, our network uses spiking neurons and operates in the temporal domain: when presented with an image, the first layer's S_1 cells, emulating V1 simple cells, detect edges with four preferred orientations and the more strongly a cell is activated the earlier it fires (intensity-to-latency conversion). A 1-WTA also ensures that at a given location only the spike corresponding to the best matching orientation is propagated. These S_1 spikes are then propagated asynchronously through the feedforward network of integrate-and-fire neurons. We only compute the first spike fired by each neuron, which leads to efficient implementations.

Note that within this time-to-first-spike framework, the maximum operation of complex cells simply consists in propagating the first spike emitted by a given group of afferents [50]. This can be done efficiently with an integrate-and-fire neuron with low threshold that has synaptic connections from all neurons in the group.

Images are presented sequentially and the resulting spike waves are propagated through to the S_2 layer, where STDP is used. We use restricted receptive fields (*i.e.* S_2 cells only integrate spikes from a $s \times s$ square neighborhood in the C_1 maps corresponding to one given processing scale) and weight sharing (*i.e.* each prototype S_2 cell is duplicated in retinotopic maps and at all scales. Of course this is not biologically plausible. In the brain, S_2-like cells probably develop their own selectivity independently, and then C_2 cells may connect to the ones with similar selectivities thanks to a trace rule [51]). Starting with a random weight matrix (size = $4 \times s \times s$) we present the first visual stimulus. Duplicated cells are all integrating the spike train in parallel, and compete with each other. If no cell reaches its threshold nothing happens and we process the next image. Otherwise for each prototype the first duplicate to reach its threshold is the winner. A 1-WTA mechanism prevents the other duplicated cells from firing. The winner thus fires and the STDP rule is triggered. The weight matrix is updated, and the change in weights is duplicated at all positions and scales. This allows the system to learn patterns despite of changes in position and size in the training examples. We also use local inhibition between different prototype cells: when a cell fires at a given position and scale, it prevents all other cells from firing later at the same scale and within an $s/2 \times s/2$ square neighborhood relative to the firing position. This competition prevents all the cells from learning the same pattern. Instead, the cell population self-organizes, each cell trying to learn a distinct pattern so as to cover the whole variability of the inputs.

If the stimuli have visual features in common (which should be the case if for example they contain similar objects), the STDP process will extract them. Fig. 8 shows an example with faces: three C_2 cells gradually become selective to face features, while at the same time, their responses become more and more rapid. Note that the background is generally not learned (at least not in priority), since backgrounds are almost always too different from one image to another for the STDP process to converge.

Importantly, the algorithm has a natural trend to learn salient regions, simply because they correspond to the earliest spikes, with the result that neurons whose receptive fields cover salient regions are likely to reach their threshold (and trigger the STDP rule) before neurons 'looking' at other regions.

In short the mechanism extracts prototypical visual patterns that are both salient and consistently present in the images. It is important to note that up to this point, the learning was fully unsupervised. No external teacher signal or previous knowledge was given to the model. For example in Fig. 8 the system had no idea it was going to see faces. The features were only learned due to statistical regularities in the dataset. However, the output of the STDP neurons can be fed into a supervised classifier, *e.g.* a RBF. We evaluated such a scheme on two Caltech datasets, one containing faces and the other motorbikes, and a distractor set containing backgrounds, all available at `www.vision.caltech.edu`. We used a standard cross-validation procedure, and previously unseen examples were correctly classified in more than 97% of the cases, using only ten C_2 cells [47].

IV. Hardware Implementation with Address Event Representation

As said above the primate visual system relies on spike arrival times to rapidly process information. However software

simulations of these mechanisms can be time consuming. Consider for example the spikes coming from a neural layer that performs a convolution on an input image, and the above-mentioned coding scheme where the most strongly activated neurons fire first. Now say you want to do a k-WTA. To select only the first k spikes one needs to do the convolution over the whole image, then sort the values, and finally discard the lowest ones. If a physical system has a response time proportional to the convolution value then the information is available as soon as the first k units have responded. This information can be transmitted to other layers for further processing, without needing to wait for later responses.

Linares' group in Sevilla, Spain is developing hardware based on this idea [52]. Specifically, they use Address Event Representation (AER), a spike-based representation technique for communicating asynchronous spikes between layers of neurons in different chips. The spikes in AER are carried as addresses of sending or receiving neurons on a digital bus. Time 'represents itself' as the asynchronous occurrence of the event. AER was first proposed in 1991 by Mead's Lab at California Institute of Technology [53], and has been used since then by a wide community of hardware engineers. As explained above, AER is ideal for WTA implementation [54].

Current devices already have some STDP-like learning capabilities. For example in [52] the so called 'CAVIAR' device was able to classify the phases of a 'toy' visual stimulus consisting of rotating disc in a purely unsupervised way. The next step will be to learn more complex visual features as in [47], useful to deal with natural images. Very relevantly it has recently been shown how the memristance nanotechnology can provide a hardware implementation of the STDP functionality [55]. This technology may be used in the next generation of AER devices.

This line of research is exciting because there is much room for improving the processing speed of biological visual systems. Indeed, biological hardware is incredibly slow: neurons cannot fire more that a few hundred spikes per second and those impulses propagate on axons between neurons with a velocity of at most a few tens of meters per second. Silicon hardware is several orders of magnitude faster. This means that a system based on biological algorithms implemented on silicon hardware could, in principle, clearly outperform animals including humans.

V. Discussion

A. Comparison to other approaches

Visual features can also be extracted using supervised learning algorithms. For example LeCun & Bengio showed how visual features in a convolutional network could be learned in a supervised manner using back-propagation [41]. Alternatively, to select only pertinent features, among a randomly selected set, for a given task, Ullman *et al.* proposed an interesting criterion based on mutual information [46]. Similarly, feature selection can be done on the basis of their likelihood ratios [56]. But supervised learning optimizes the features set for a given class (*e.g.* faces, cars, *etc.*).

In this paper, we focused on unsupervised visual feature learning. The features then reflect the statistics of the environment [57]. Specifically, they correspond to patterns that are consistently present in the inputs. This clustering removes redundancy [58], which is desirable to a certain extent, although other criteria such as the sparseness of the resulting code should also be taken into account [59], [60]. Then the output of the feature detector can be fed into a supervised learning algorithm (*e.g.* RBF, Support Vector Machine *etc.*). Note that this kind of hybrid scheme has been found to learn much faster than a two-layer backpropagation network, because the credit assignment problem is facilitated when features are kept fixed [61], [62].

Many other approaches have been proposed for unsupervised learning of visual features. Some authors use random crops from natural images (*e.g.* [48]). This works well because it can exhaustively sample the input space [63], but it is costly since redundancy is very high between features, and many features are irrelevant for most (if not all) of the tasks.

Other approaches use Independent Component Analysis (ICA) or derivatives to remove as much redundancy as possible in the input, for *e.g.* [64]–[67]. However, the independence assumption of ICA is ill suited for learning part-based representations because various parts that are likely to occur together would end up in a single holistic representation. The STDP-based algorithm proposed here learns each part independently, which is clearly a little redundant but leads to better robustness to occlusion and better generalization performance.

Principal Component Analysis (PCA) is another popular approach (*e.g.* [68], [69]), but it leads to features that lack intuitive meaning because of complex cancellations between positive and negative numbers while combining them [70].

The STDP-based learning algorithm is much closer to non-negative matrix factorization [70] or non-negative sparse coding [71]: within both approaches objects are represented as (positive) sums of their parts, and the parts are learned by detecting consistently co-active input units.

B. Next steps

One of the main limitation of the studies presented in this paper is that they considered only static images, and propagated them one by one. A next step could be to use videos. A recent retinal model developed at the INRIA [72] can be used to convert videos into spikes, which would then be fed to the network. STDP is known to be able to detect consistent spike patterns even when embedded in continuous activity [73], [74], so selectivity to prototypical visual features should thus still emerge. Besides, spiking neurons and rank order schemes are also known to be capable of motion detection and integration [75], leading to multilayered networks that can robustly classify human actions (*e.g.* walk, jump *etc.*) in natural videos [76]. But

whether STDP can learn motion features is currently unclear, and subsequent work will address this issue.

The second main limitation of the studies presented here is the restriction to feedforward connections. The main justification for these feedforward-only models is that, as explained above, the primate visual system seems to have a fast recognition mode in which feedback is probably largely inactive. However normal vision is an ongoing process, in which the massive amount of feedback connections observed in the brain [77], [78] certainly have important functional roles.

One of them maybe to generate self-sustained oscillations [79]. It has been recently demonstrated that a common oscillatory drive for a group of neurons can reliably format the pattern of spike times – through an activation-to-phase conversion [80] – so that repeating activation patterns can be easily detected and learned by a downstream neuron equipped with STDP, and then recognized in just one oscillation cycle [81]. But how STDP interacts with self-sustained oscillations, as opposed to an external oscillatory drive, and the implications for visual processing are currently unclear. Again, subsequent work will address theses issues.

While omnipresent in biological visual systems oscillations have received little attention from the computer vision community. Yet they may have a lot of desirable effects such as organizing perception in discrete chunks. Alternatively, saccades may have a similar effect [81], [82].

ACKNOWLEDGMENT

This work was supported in part by the Fyssen Foundation, CNRS, STREP Decisions-in-Motion (IST-027198), ANR (Projects Natstats and Hearing in Time), and SpikeNet Technology SARL.

REFERENCES

[1] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system," *Nature*, vol. 381, no. 6582, pp. 520–2, 1996.

[2] M. Fabre-Thorpe, G. Richard, and S. J. Thorpe, "Rapid categorization of natural images by rhesus monkeys," *Neuroreport*, vol. 9, no. 2, pp. 303–8, 1998.

[3] G. A. Rousselet, M. Fabre-Thorpe, and S. J. Thorpe, "Parallel processing in high-level categorization of natural images," *Nat Neurosci*, vol. 5, no. 7, pp. 629–30., 2002.

[4] N. Bacon-Mace, M. J. Mace, M. Fabre-Thorpe, and S. J. Thorpe, "The time course of visual processing: Backward masking and natural scene categorisation," *Vision Res*, vol. 45, no. 11, pp. 1459–69, 2005.

[5] H. Kirchner and S. Thorpe, "Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited," *Vision Research*, vol. 46, no. 11, pp. 1762–1776, 2006.

[6] T. Serre, A. Oliva, and T. Poggio, "A feedforward architecture accounts for rapid categorization," *Proc. Nat. Acad. Sci. USA*, vol. 104, no. 15, 2007.

[7] P. Girard, C. Jouffrais, and C. H. Kirchner, "Ultra-rapid categorisation in non-human primates." *Anim Cogn*, vol. 11, no. 3, pp. 485–493, Jul 2008.

[8] M. Oram and D. Perrett, "Time course of neural responses discriminating different views of the face and head." *J Neurophysiol*, vol. 68, no. 1, pp. 70–84, 1992.

[9] C. Keysers, D. K. Xiao, P. Földiák, and D. I. Perrett, "The speed of sight," *J. Cogn. Neurosci.*, vol. 13, pp. 90–101, 2001.

[10] C. Hung, G. Kreiman, T. Poggio, and J. DiCarlo, "Fast readout of object identity from macaque inferior temporal cortex." *Science*, vol. 310, no. 5749, pp. 863–866, 2005.

[11] H. Liu, Y. Agam, J. R. Madsen, and G. Kreiman, "Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex." *Neuron*, vol. 62, no. 2, pp. 281–290, Apr 2009.

[12] S. Thorpe and M. Imbert, "Biological constraints on connectionist modelling," in *Connectionism in perspective*. Amsterdam: Elsevier, 1989, pp. 63–92.

[13] S. Thorpe and J. Gautrais, "Rank order coding," in *Computational Neuroscience : Trends in Research*, J. M. Bower, Ed. New York: Plenum Press, 1998, pp. 113–118.

[14] D. G. Albrecht, W. S. Geisler, R. A. Frazor, and A. M. Crane, "Visual cortex neurons of monkeys and cats: temporal dynamics of the contrast response function," *J Neurophysiol*, vol. 88, no. 2, pp. 888–913., 2002.

[15] T. Gawne, T. Kjaer, and B. Richmond, "Latency : another potential code for feature binding in striate cortex." *J Neurophysiol*, vol. 76, no. 2, pp. 1356–1360, 1996.

[16] S. Celebrini, S. Thorpe, Y. Trotter, and M. Imbert, "Dynamics of orientation coding in area V1 of the awake primate." *Vis Neurosci*, vol. 10, no. 5, pp. 811–825, 1993.

[17] S. Thorpe, "Spike arrival times: A highly efficient coding scheme for neural networks," in *Parallel processing in neural systems and computers*, R. Eckmiller, G. Hartmann, and G. Hauske, Eds. Elsevier, 1990, pp. 91–94.

[18] S. Elias and S. Grossberg, "Pattern formation, contrast control, and oscillations in the short term memory of shunting on-center off-surround networks," *Biol. Cyb.*, vol. 20, pp. 69–98, 1975.

[19] S. Amari and M. Arbib, *Systems Neuroscience*. Academic Press (San Diego), 1977, ch. Competition and cooperation in neural nets., pp. 119–165.

[20] A. L. Yuille and N. M. Grzywacz, "A winner-take-all mechanism based on presynaptic inhibition feedback." *Neural Comp.*, vol. 1, no. 3, pp. 334–347, 1989.

[21] R. Coultrip, R. Granger, and G. Lynch, "A cortical model of winner-take-all competition via lateral inhibition," *Neural Networks*, vol. 5, pp. 47–54, 1992.

[22] A. J. Yu, M. A. Giese, and T. Poggio, "Biophysiologically plausible implementations of the maximum operation," *Neural Comp.*, vol. 14, no. 12, pp. 2857–2881, 2002.

[23] U. Knoblich, J. Bouvrie, and T. Poggio, "Biophysical models of neural computation: Max and tuning circuits," CBCL Paper, MIT, Tech. Rep., 2007.

[24] M. Kouh and T. Poggio, "A canonical neural circuit for cortical nonlinear operations." *Neural Comput*, vol. 20, no. 6, pp. 1427–1451, Jun 2008.

[25] R. VanRullen and S. Thorpe, "Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex." *Neural Comput*, vol. 13, no. 6, pp. 1255–1283, 2001.

[26] R. Guyonneau, R. VanRullen, and S. Thorpe, "Temporal codes and sparse representations: a key to understanding rapid processing in the visual system." *J Physiol Paris*, vol. 98, no. 4-6, pp. 487–497, 2004.

[27] N. Caporale and Y. Dan, "Spike timing-dependent plasticity: a hebbian learning rule." *Annu Rev Neurosci*, vol. 31, pp. 25–46, 2008.

[28] D. O. Hebb, *The organization of behavior*. Wiley, New York, 1949.

[29] G. Bi and M. M. Poo, "Synaptic modification by correlated activity : Hebb's postulate revisited," *Ann Rev Neurosci*, vol. 24, pp. 139–166, 2001.

[30] M. C. van Rossum, G. Q. Bi, and G. G. Turrigiano, "Stable hebbian learning from spike timing-dependent plasticity." *J Neurosci*, vol. 20, no. 23, pp. 8812–8821, Dec 2000.

[31] R. Gütig, R. Aharonov, S. Rotter, and H. Sompolinsky, "Learning input correlations through nonlinear temporally asymmetric hebbian plasticity." *J Neurosci*, vol. 23, no. 9, pp. 3697–3714, May 2003.

[32] A. N. Burkitt, H. Meffin, and D. B. Grayden, "Spike-timing-dependent plasticity: the relationship to rate-based learning for models with weight dynamics determined by a stable fixed point." *Neural Comput*, vol. 16, no. 5, pp. 885–940, May 2004.

[33] S. Song, K. Miller, and L. Abbott, "Competitive hebbian learning through spike-timing-dependent synaptic plasticity." *Nat Neurosci*, vol. 3, no. 9, pp. 919–926, 2000.

[34] W. Gerstner and W. Kistler, *Spiking Neuron Models*. Cambridge Univ. Press, 2002.

[35] R. Guyonneau, R. VanRullen, and S. Thorpe, "Neurons tune to the earliest spikes through STDP." *Neural Comput*, vol. 17, no. 4, pp. 859–879, 2005.

[36] U. Weidenbacher and H. Neumann, "Unsupervised learning of head pose through spike-timing dependent plasticity," in *Perception in Multimodal Dialogue Systems*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2008, vol. 5078/2008, pp. 123–131.

[37] R. Guyonneau, "Codage par latence et stdp: des stratégies temporelles pour expliquer le traitement visuel rapide," Ph.D. dissertation, Université Toulouse III - Paul Sabatier, 2006.

[38] P. Phillips, H. Moon, P. Rauss, and S. Rizvi, "The FERET evaluation methodology for face recognition algorithms," *IEEE Trans Pattern Anal Mach Intell*, vol. 22, no. 10, p. 10901104, 2000.

[39] K. Fukushima, "Neocognitron : a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position." *Biol Cybern*, vol. 36, no. 4, pp. 193–202, 1980.

[40] G. Wallis and E. Rolls, "Invariant face and object recognition in the visual system." *Prog Neurobiol*, vol. 51, no. 2, pp. 167–194, 1997.

[41] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed. Cambridge, MA: MIT Press, 1998, pp. 255–258.

[42] R. VanRullen, J. Gautrais, A. Delorme, and S. Thorpe, "Face processing using one spike per neurone." *Biosystems*, vol. 48, no. 1-3, pp. 229–239, 1998.

[43] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex." *Nat Neurosci*, vol. 2, no. 11, pp. 1019–1025, 1999.

[44] E. Rolls and T. Milward, "A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures." *Neural Comput*, vol. 12, no. 11, pp. 2547–2572, 2000.

[45] S. Stringer and E. Rolls, "Position invariant recognition in the visual system with cluttered environments." *Neural Netw*, vol. 13, no. 3, pp. 305–315, 2000.

[46] S. Ullman, M. Vidal-Naquet, and E. Sali, "Visual features of intermediate complexity and their use in classification." *Nat Neurosci*, vol. 5, no. 7, pp. 682–687, 2002.

[47] T. Masquelier and S. J. Thorpe, "Unsupervised learning of visual features through spike timing dependent plasticity." *PLoS Comput Biol*, vol. 3, no. 2, p. e31, Feb 2007.

[48] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Trans Pattern Anal Mach Intell*, vol. 29, no. 3, pp. 411–426, 2007.

[49] A. Delorme, L. Perrinet, S. Thorpe, and M. Samuelides, "Networks of integrate-and-fire neurons using rank order coding B: Spike timing dependent plasticity and emergence of orientation selectivity," *Neurocomputing*, vol. 38-40, pp. 539–545, 2001.

[50] G. Rousselet, S. Thorpe, and M. Fabre-Thorpe, "Taking the max from neuronal responses." *Trends Cogn Sci*, vol. 7, no. 3, pp. 99–102, 2003.

[51] T. Masquelier, T. Serre, S. Thorpe, and T. Poggio, "Learning complex cell invariance from natural videos: a plausibility proof." *Massachusetts Institute of Technology*, vol. CBCL Paper #269 / MIT-CSAIL-TR #2007-060, 2007.

[52] R. Serrano-Gotarredona, M. Oster, P. Lichtsteiner, A. Linares-Barranco, R. Paz-Vicente, F. Gomez-Rodriguez, L. Camunas-Mesa, R. Berner, M. Rivas-Perez, T. Delbruck, S.-C. Liu, R. Douglas, P. Hafliger, G. Jimenez-Moreno, A. Ballcels, T. Serrano-Gotarredona, A. Acosta-Jimenez, and B. Linares-Barranco, "Caviar: A 45k neuron, 5m synapse, 12g connects/s aer hardware sensory–processing–learning–actuating system for high-speed visual object recognition and tracking," *IEEE TRANSACTIONS ON NEURAL NETWORKS*, vol. 20, no. 9, pp. 1417–1438, 2009.

[53] M. Sivilotti, "Wiring considerations in analog vlsi systems with application to field-programmable networks," Ph.D. dissertation, Comput. Sci. Div., California Inst. Technol., Pasadena, CA, 1991.

[54] Z. Kalayjian and A. G. Andreou, "Asynchronous communication of 2d motion information using winner-takes-all arbitration," *Int. J. Anal. Integr. Circuits Signal Process.*, vol. 13, no. 1-2, pp. 103–109, 1997.

[55] B. Linares-Barranco and T. Serrano-Gotarredona, "Memristance can explain spike-time-dependent-plasticity in neural synapses," *Nature Precedings*, 2009.

[56] G. Dorko and C. Schmid, "Selection of scale-invariant parts for object class recognition," in *Proc. Ninth IEEE International Conference on Computer Vision*, 2003, pp. 634–639 vol.1.

[57] G. E. Hinton and T. J. Sejnowski, Eds., *Unsupervised learning: foundations of neural computation*. The MIT Press, 1999.

[58] H. Barlow, *Sensory Communication*, wa rosenblith ed. Cambridge, MA: MIT Press, 1961, ch. Possible principles underlying the transformation of sensory messages, pp. 217–234.

[59] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells." *J Opt Soc Am A*, vol. 4, no. 12, pp. 2379–2394, Dec 1987.

[60] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.

[61] E. Rolls and G. Deco, *Computational neuroscience of vision*. Oxford University Press, 2002.

[62] M. Ranzato, F. J. Huang, Y.-L. Boureau, and Y. LeCun, "Unsupervised learning of invariant feature hierarchies with applications to object recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition CVPR '07*, 2007, pp. 1–8.

[63] F. Jurie and B. Triggs, "Creating efficient codebooks for visual recognition," in *Proc. Tenth IEEE International Conference on Computer Vision ICCV 2005*, vol. 1, 2005, pp. 604–610 Vol. 1.

[64] A. J. Bell and T. J. Sejnowski, "The "independent components" of natural scenes are edge filters." *Vision Res*, vol. 37, no. 23, pp. 3327–3338, Dec 1997.

[65] J. H. van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex." *Proc Biol Sci*, vol. 265, no. 1394, pp. 359–366, 1998.

[66] P. O. Hoyer and A. Hyvarinen, "A multi-layer sparse coding network learns contour coding from natural images," *Vision Res*, vol. 42, no. 12, pp. 1593–605., 2002.

[67] Y. Karklin and M. S. Lewicki, "Learning higher-order structures in natural images," *Network*, vol. 14, no. 3, pp. 483–99., 2003.

[68] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 696–710, 1997.

[69] K. Murphy, A. Torralba, and W. T. Freeman, "Using the forest to see the trees: A graphical model relating features, objects, and scenes," in *Advances in Neural Information Processing Systems 16*, S. Thrun, L. Saul, and B. Schölkopf, Eds. MIT Press, 2004.

[70] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization." *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[71] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding." *Journal of Machine Learning Research (in press)*, 2010.

[72] A. Wohrer and P. Kornprobst, "Virtual retina: a biological retina model and simulator, with contrast gain control." *J Comput Neurosci*, vol. 26, no. 2, pp. 219–249, Apr 2009.

[73] T. Masquelier, R. Guyonneau, and S. J. Thorpe, "Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains." *PLoS ONE*, vol. 3, no. 1, p. e1377, 2008.

[74] ——, "Competitive STDP-based spike pattern learning." *Neural Comput*, vol. 21, no. 5, pp. 1259–1276, May 2009.

[75] C. Beck, S. Thorpe, and H. Neumann, "Neural rank-order coding with spiking neurons for cortical motion detection and integration," in *Proc. Int'l. Conf. on Cognitive Systems, CogSys 2008, Karlsruhe, Germany*, April 2008.

[76] M. J. Escobar, G. S. Masson, and T. V. P. Kornprobst, "Action recognition using a bio-inspired feedforward spiking network," *Int J Comput Vis*, vol. 82, p. 284301, 2009.

[77] D. J. Felleman and D. C. Van Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cereb Cortex*, vol. 1, no. 1, pp. 1–47, 1991.

[78] R. J. Douglas and K. A. Martin, "Neuronal circuits of the neocortex," *Annu Rev Neurosci*, vol. 27, pp. 419–51, 2004.

[79] C. M. Gray and W. Singer, "Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex." *Proc Natl Acad Sci U S A*, vol. 86, no. 5, pp. 1698–1702, Mar 1989.

[80] J. Hopfield, "Pattern recognition computation using action potential timing for stimulus representation." *Nature*, vol. 376, no. 6535, pp. 33–36, 1995.

[81] T. Masquelier, E. Hugues, G. Deco, and S. J. Thorpe, "Oscillations, phase-of-firing coding, and spike timing-dependent plasticity: an efficient learning scheme." *J Neurosci*, vol. 29, no. 43, pp. 13 484–13 493, Oct 2009.

[82] N. Uchida, A. Kepecs, and Z. F. Mainen, "Seeing at a glance, smelling in a whiff: rapid forms of perceptual decision making," *Nat Rev Neurosci*, vol. 7, no. 6, pp. 485–491, 2006.