Available online at www.sciencedirect.com
SCIENCE DIRECT
NEUROCOMPUTING
Neurocomputing 58–60 (2004) 857–864
www.elsevier.com/locate/neucom

SpikeNet: real-time visual processing with one spike per neuron

Simon J. Thorpe[a,b,*], Rudy Guyonneau[a,b], Nicolas Guilbaud[a,b], Jong-Mo Allegraud[a,b], Rufin VanRullen[a,b]

[a] *Centre de Recherche Cerveau & Cognition UMR 5549, Univ. Paul Sabatier, 133, route de Narbonne, Toulouse 31062, France*
[b] *SpikeNet Technology S.A.R.L., Ave de Castelnaudary, Revel 31250, France*

Abstract

SpikeNet is an image-processing system that uses very large-scale networks of asynchronously firing neurons. The latest version allows very efficient object identification in real-time using a video input, and although this specific implementation is designed to run on standard computer hardware, there are a number of clear implications for computational neuroscience. Specifically, SpikeNet demonstrates the plausibility of visual processing based on a single feed-forward pass and very sparse levels of firing. Above all, it is one of the very few models compatible with the severe temporal constraints imposed by experimental data on processing speed in the visual system.

Keywords: Spiking neurons; Visual processing; Temporal coding; Shunting inhibition

1. Introduction

The speed with which the brain processes flashed images provides a major challenge for computational neuroscience. For example, we recently reported that when two previously unseen images are flashed to the left and right of a fixation point, subjects can make reliable saccades to the side where the image contains an animal in as little as 130 ms [6]. Given the anatomy of the visual and oculomotor pathways, this implies that at each stage of the processing chain, decisions need to be taken very quickly,

* Corresponding author. Centre de Recherche Cerveau & Cognition UMR 5549, Univ. Paul Sabatier, 133, route de Narbonne, Toulouse 31062, France. Tel.: +33-5-62-17-28-03; fax: +33-5-62-17-28-09.
E-mail address: thorpe@cerco.ups-tlse.fr (S.J. Thorpe).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2004.01.138

possibly in less than 10 ms. This poses a major problem for conventional processing models based on firing rate, since 10 ms is too short a time to accurately determine the firing rate of individual cells.

For some time, we have been exploring the possibility that the order in which neurons fire could be used as a code even when only very few cells have time to emit a spike [5,8]. Although it is difficult to prove such hypotheses experimentally, computer simulation provides a way of determining whether or not the approach is feasible. Here we present results obtained with SpikeNet, an image processing system based on large arrays of asynchronously spiking neurons. While not intended to be a highly realistic model of visual processing, the basic features of SpikeNet are all ones that are compatible with the basic properties of neurons. As a result, the fact that the system is capable of real-time object recognition and localization demonstrates the viability of such an approach.

2. Basic mechanisms—from temporal order to selectivity

SpikeNet is based on the idea that the order of firing within a population of cells can be used to encode information. As illustrated in Fig. 1, an activation profile, produced for example by an image flashed on the retina will lead to a wave of spikes in which the earliest firing neurons will generally correspond to the most strongly activated neurons. The idea that order of firing can be used to encode information follows naturally from the fact that even the simplest integrate-and-fire neuron model will reach threshold more rapidly when the pattern of inputs matches its selectivity. Evidence in support of such a view came from a theoretical study by VanRullen and Thorpe [8] who used a simple model of the receptive field properties of ganglion cells in the retina to demonstrate that when the order of firing of retinal ganglion cells is taken into account, the input image can be reconstructed with sufficient accuracy to allow identification of many stimuli when a few as 0.5–1.0% of the cells have fired.

Fig. 2 shows how the addition of a feed-forward shunting inhibition circuit can be used to produce neurons that respond selectively as a function of the order of firing. The inhibitory unit S receives equally effective excitation from all the input units, and progressively desensitizes the two target neurons as a function of the number of inputs

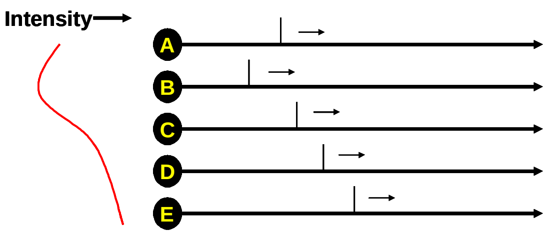

Fig. 1. In response to an intensity profile, the neurons A–E will generate a wave of spikes in which the earliest firing units correspond to the most strongly activated cells.

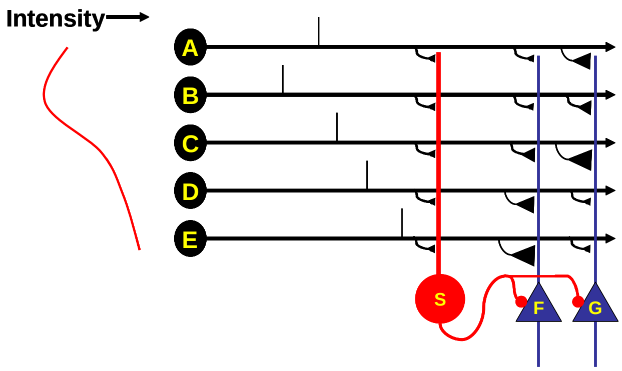

Fig. 2. Units F and G receive excitatory synapses from the input units with variable weights. Unit S receives fixed excitatory synapses from all the inputs and generates shunting inhibition that progressively desensitizes units F and G as more and more of the inputs has fired.

that have fired. Thus, while the first inputs to fire are fully effective, inputs that fire later produce less and less activation. Under such conditions, the total amount of excitation produced in units F and G will depend on how well the order of firing matches the pattern of weights from the input units. Maximal activation is produced when the order of firing of the inputs matches the set of synaptic weights. For example, since unit F has relatively strong weights from inputs C, D and E, it will respond well in response to the current input pattern.

Interestingly, recent experimental work on the properties of fast spiking interneurons in the cortex has provided evidence that supports the existence of such a mechanism. For example, in the somatosensory cortex, Swadlow and Gusev have shown the fast spiking inhibitory interneurons have very small somas and can react very rapidly to the activation of their inputs. Their very short duration action potentials allow them to fire at very high rates up to 600 spikes per second, way above the values seen for the vast majority of cortical neurons. Furthermore, they receive strong but very nonselective inputs from the thalamus with the result that they show essentially no stimulus selectivity [7]. Stimulus selectivity of the inhibitory units will also be reduced by the existence of electrical coupling between these cells [4], which will tend to make the entire population fire together. Finally, there is now good evidence that the inhibitory units produce shunting inhibition in the target neurons. Indeed, the shunting inhibition develops very rapidly following the presentation of a visual stimulus, leading to a three fold increase in soma conductance [1].

Feed-forward inhibition is not the only mechanism that may be involved. Fig. 3 shows how the addition of a feedback inhibitory circuit can also be computationally useful. In this case, by adjusting the threshold for activation of the inhibition, one can prevent more than a certain number of cells in the input layer from generating a spike. If the inhibition is very strong, such a circuit will effectively perform a Winner-Take-All operation on the inputs, allowing only the most strongly activated input to fire. But with a higher threshold, the operation is more like a k-Winner-Take-All operation.

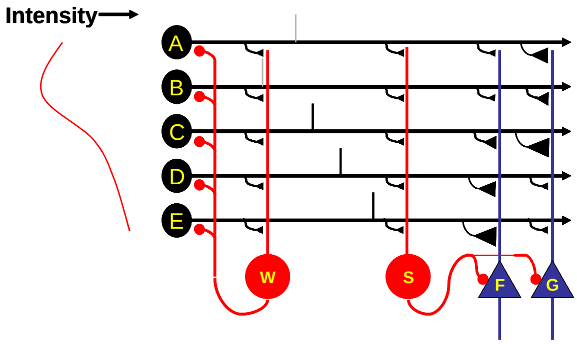

Fig. 3. Addition of a feed-back inhibitory circuit (W) allows the implementation of a Winner-Take-All operation that prevents more than a limited number of cells in the input layer from firing. Here, the threshold for triggering the feed-back inhibition has been fixed at 3 so that only the first 3 spikes are able to pass.

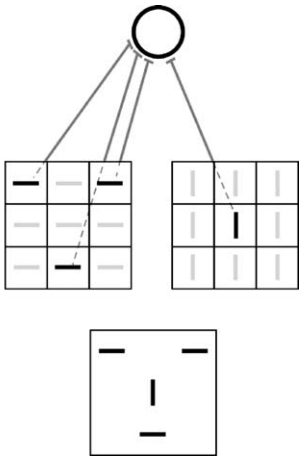

Fig. 4. A very simple illustration of how selectivity for fairly high level representations (here a face-like pattern) could be generated by using a limited number of connections and low levels of activation. The output neuron receives connections from two 3×3 arrays of feature detectors, tuned to vertical and horizontal orientation. If, as a result of learning, only four of the input neurons have strong connections, and if when a new input image is presented only the first four input units are allowed to fire, the output neuron will only receive maximal activation when all four features (and no others) are present.

Together, rapidly acting feed-forward and feedback inhibitory circuits provide mechanisms that can (i) allow the percentage of active cells in the sensory pathways to be controlled, and (ii) make neurons in the next layer sensitive to the order in which the inputs fire. How might such mechanisms allow neurons at later stages to respond selectively to particular inputs?

Fig. 4 illustrates the very simple case of a neuron receiving from two 3×3 unit arrays of orientation selective feature detectors. Suppose that initially, the unit receives weak connections from all 18 input units, but that as a result of learning, all the weights are concentrated on just four of the inputs—3 horizontally tuned units, and one vertically tuned unit. If we now use the sort of inhibitory circuits described in the preceding paragraphs to limit the number of active units in the input arrays to 4, it should be clear that the chances of all four units matching the input pattern of the cell will be very low. If we fix the threshold for firing in the output neuron at 4 active inputs, only one of the 3060 possible ways of activating four units in the input array will generate a spike in the output neuron. This illustrates that even a very simple mechanism can be selective for something that can be quite high level because in the case illustrated here, the features correspond to something quite specific, namely, a face.

3. Simulations with SpikeNet

Could this sort of simple mechanism be used in the brain? And could it go some way to explaining the extraordinary processing power of the visual system? One approach is to use computer simulations to see whether a system based on such principles is able to perform tasks requiring visual recognition. SpikeNet is a simulation package that aims to do just this. There are actually two different versions of SpikeNet. The original version was developed by Arnaud Delorme during his doctoral thesis [2], whereas a more recent version has been developed for image processing. The source code of Delorme's original version can now be downloaded under GNU licence from his website and is described in detail in a recent article [3]. The more recent version is a commercial product developed by SpikeNet Technology SARL under licence from the CNRS, but a demonstration version can be downloaded from the company's website at http://www.spikenet-technology.com. While for the commercial version, priority has been given to providing a software package capable of reliable real-time image processing in real-world situations, both versions share the same underlying computational principles. In particular, both versions test the idea that high level visual processing tasks can be performed under conditions where each neuron in the system only gets to fire at most one spike, and where the percentage of neurons that actually emit spikes is kept to a strict minimum. Obviously, in the real nervous system, it is (at least for the foreseeable future) impossible to prevent neurons from firing multiple spikes. As a result, it is unlikely that it will be possible to *prove* that the nervous system can perform high-level visual tasks with only one spike per neuron. However, by building a synthetic system such as SpikeNet in which multiple spiking is prevented we can ask just how much visual processing can be achieved under such conditions.

Fig. 5 is a screen-shot of an application based on the SpikeNet kernel in which a single visual model has been used to locate both eyes in a static image of a cat. In this case only one model is used, but there are no limits on the number of different "models" that can be implemented. The basic architecture used in these simulations is illustrated in Fig. 6. A pre-processing stage corresponding roughly to the retina and V1 and the resulting representation contains a set of orientation-selective "maps" each roughly corresponding to "simple" type orientation selective neurons. These neurons reach threshold and fire at a latency that depends on the strength of the input. Thus,

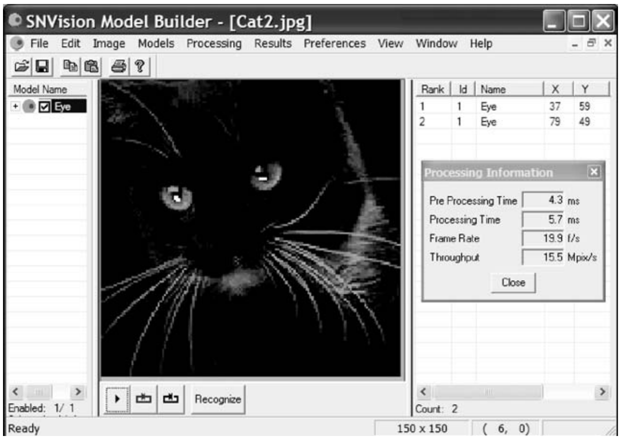

Fig. 5. Interface of SpikeNet. The image to be processed is shown in the central panel. To the left is a list of the different models to be recognized. In this case, for simplicity, we show a single model trained with the eye. On the right, the panel shows that two separate regions of activation were found, corresponding to the two eyes. The pixels where the recognition layers units have fired are shown by the white points within the eyes. The Processing Information window shows that the total processing time for analyzing this 150×150 pixel image was 5.7 ms on a 2 GHz Pentium 4 based machine.

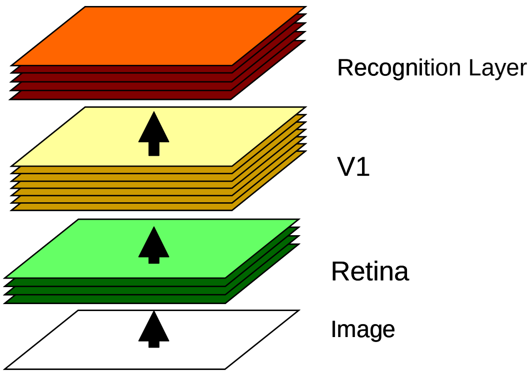

Fig. 6. The current version of SpikeNet generally uses a very simple architecture in which arrays of units in the recognition layer receive direct inputs from the equivalent of V1.

if a high contrast vertical edge is present at a particular point in the input image, the neuron with a vertically tuned receptive field at the appropriate location will be one of the first to reach threshold and fire. In this way, the order of firing within V1 contains information about the contours present in the image. When a new visual form (or model) is learnt, an array of neurons with the same dimensions as the input image is created, with one unit for each pixel. All of these neurons share the same pattern of weights, determined by a learning algorithm which fixes high weights with the earliest firing inputs and low or zero weights for the others. This allows us to recognize (and locate) the same visual form anywhere within the image which effectively provides translation invariance, although the cost in terms of the number of neurons required is extremely high.

Clearly such an arrangement is very different to the one used by the visual system in which several layers of processing are interleaved between V1 and the equivalent of the "recognition layer" which is presumably located in something like inferotemporal cortex. As one progresses through extrastriate areas such as V2 and V4, receptive field sizes increase until inferotemporal cortex where receptive fields can include much of the visual field. It is likely that this is a way of obtaining position invariance that reduces the total number of neurons required. In SpikeNet, we can effectively recognize hundreds or even thousands of different visual forms by creating a new array of recognition units for each new object. But the cost in terms of the number of neurons involved would be astronomical because one would need one unit for each point in the image. With images containing up to a million different pixels, it would be totally prohibitive to adopt such an approach in the human visual system. On the other hand, SpikeNet does not suffer from the binding problem since object identity and position are explicitly coded. Indeed, the fact that the system can report the number and location of each object in the scene is a distinct advantage.

4. Conclusions and perspectives

We make no claim that the current architecture used in SpikeNet is realistic. However, the main point that we wish to make concerns the efficiency of the underlying algorithm. By restricting the number of neurons that fire and using only the first 1–2% of cells that fire, highly selective visual responses can be produced very rapidly. These results allow us to draw the following tentative conclusions.

Single spike coding is viable: The first point is that sophisticated visual processing with just one spike per neuron is clearly possible, despite the fact that with only one spike, traditional coding schemes based on determining the firing rates of individual neurons are ruled out.

Pure feed-forward mechanisms are computationally powerful: Although SpikeNet can include both horizontal and feedback connection patterns, the current version does not use them. Despite this, accurate identification is possible even with noisy images and at low contrasts. Clearly, the initial feed-forward wave of processing is capable of considerably more than is conventionally assumed.

Sparse coding is very efficient: One of the main reasons for the speed of SpikeNet lies in its very sparse coding scheme. Typically, we have found that only 1–2% of neurons in any given processing stage need to fire in order to allow identification. The key is to use a coding scheme in which the most strongly activated neurons fire first (rank order coding) since this guarantees that decisions are made as quickly as possible.

Image segmentation is not required for high level identification: One of the most striking features of SpikeNet is that there is nothing even remotely like image segmentation going on. Everything is done by using large numbers of neurons tuned to diagnostic combinations of features that will fire as soon as there is enough evidence to allow activation. It could be that the traditional view that the first step in processing requires scene segmentation is a major error, and that intelligent segmentation involves feedback that occurs only once the initial feed-forward pass has been completed.

The processing architectures used by SpikeNet are still a long way from those used by biological vision and future work will be aimed at reducing the gap. For example, SpikeNet does not have the equivalent of separate ventral and dorsal pathways specialized for object identification and localization. Instead, there is a retino-topically organized map of neurons for each object or feature constellation that needs to be identified. For applications, this is actually quite useful, because the system automatically provides the *xy* coordinates of each identified object (unlike object selective neurons in inferotemporal cortex that have only limited spatial selectivity). However, there is a very high cost in terms of the number of neurons required. Future versions will try to use a more biologically realistic strategy that almost certainly will allow a major reduction in the number of neurons required. Nevertheless, the biological reverse engineering approach used in SpikeNet has already proved remarkably successful and a number of important computational issues have already been addressed using this sort of approach.

References

[1] L.J. Borg-Graham, C. Monier, Y. Fregnac, Visual input evokes transient and strong shunting inhibition in visual cortical neurons, Nature 393 (6683) (1998) 369–373.

[2] A. Delorme, J. Gautrais, R. Van Rullen, S.J. Thorpe, SpikeNET: a simulator for modeling large networks of integrate and fire neurons, NeuroComputing 26–27 (1999) 989–996.

[3] A. Delorme, S.J. Thorpe, SpikeNET: an event-driven simulation package for modelling large networks of spiking neurons, Network: Comput. Neural Systems 14 (4) (2003) 613–628.

[4] M. Galarreta, S. Hestrin, Electrical synapses between GABA-releasing interneurons, Nature Rev. Neurosci. 2 (6) (2001) 425–433.

[5] J. Gautrais, S. Thorpe, Rate coding versus temporal order coding: a theoretical approach, Biosystems 48 (1–3) (1998) 57–65.

[6] H. Kirchner, N. Bacon, S.J. Thorpe, In which of two scenes is the animal? Ultra-rapid visual processing demonstrated with saccadic eye movements, Perception 32 (suppl) (2003) 170.

[7] H.A. Swadlow, A.G. Gusev, Receptive-field construction in cortical inhibitory interneurons, Nat. Neurosci. 5 (5) (2002) 403–404.

[8] R. VanRullen, S. Thorpe, Rate coding vs. temporal order coding: what the retinal ganglion cells tell the visual cortex, Neural Comput. 13 (6) (2001) 1255–1283.