ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 79-84

How can the visual system process a natural scene in under 150 ms? On the role of asynchronous spike propagation.

Simon Thorpe & Jacques Gautrais

Centre de Recherche Cerveau & Cognition

133, route de Narbonne, 31062, Toulouse, France

thorpe@cerco.ups-tlse.fr gautrais@cerco.ups-tlse.fr

Abstract

The human visual system can process previously unseen natural scenes in under 150 ms, even with extrafoveal viewing. Such data impose serious temporal constraints on the way in which information is processed. We argue that one of the keys to understanding this remarkable efficiency lies in the fact that biological vision makes use of asynchronous spike propagation, a feature absent in the vast majority of artificial neural networks. We describe our recent work that has explored the computational advantages of such asynchronous processing.

1. Introduction - Image processing in humans

Research on artificial neural networks (ANNs) has progressed rapidly in the last decade or so. However, even the most sophisticated ANNs have great difficulty in performing tasks that humans find trivial. Take for example our seemingly effortless ability to interpret and categorize natural visual scenes. We have examined this ability in a recent series of experiments that used a commercially available set of thousands of colour photographs (Thorpe, Fize & Marlot, 1996). The photographs were flashed on the screen of a computer display for just 20 ms, and subjects had to indicate (by releasing a button) whether the image contained an animal. Such a task makes extremely severe demands on the visual system. First, none of the images had been seen previously by the subjects, thus preventing the use of strategies specific to particular images. Second, the subjects had no idea about what particular sort of animal to look for, since the targets included a wide range of animals in their natural environments (birds, fish, mammals, reptiles etc.). Furthermore, they had no way of predicting the position, the size of the target, the lighting conditions or even the number of animals present. Despite this, performance was truly remarkable. Accuracy averaged 94%, with one subject achieving 98% correct, and median reaction times averaged 445 ms but could be as short as 382 ms. In addition, we were able to show using simultaneously recorded Event-Related Potentials (ERPs) that the visual processing necessary for performing this task could be achieved in roughly 150 ms or less.

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 79-84

In another series of experiments (Thorpe, Fabre-Thorpe & Richard, 1996) we reported that similarly rapid and accurate categorization was possible even for images that were not presented in central vision. Subjects performed the same task as in the previous study, except that the images were presented at random at one of three positions, either centrally, or at an eccentricity of 3.6° to the left or right of the fixation point. As in the previous study, performance with centrally presented images was extremely good (95.7% correct; mean reaction time : 453ms). Surprisingly, however, performance was only slightly less good with laterally presented stimuli (92.5% correct; mean reaction time : 465ms). Such small effects of extrafoveal presentations are surprising given the fact that acuity drops rapidly as retinal eccentricity is increased. Furthermore, the differential ERP response described previously on no-go trials had a similar latency with central and lateral presentations, implying that there is effectively no time penalty for processing laterally presented stimuli, even when the location of the presentation is totally unpredictable.

2. Keys to understanding biological visual processing

The remarkably efficiency with which the human visual system can process natural scenes is clearly something which artificial systems are currently unable to reproduce. Where does this extraordinary processing power come from?

2.1 The importance of parallelism

Clearly, one of the keys to understanding this efficiency lies in the use of massively parallel processing. In primates, retinal information is transmitted to the cortex by over a million fibres in each optic nerve, and the first cortical area (V1) contains more than 350 million neurons. Subsequent processing is performed in a number of hierarchically organized processing stages. Thus, to reach the inferotemporal cortex, thought to be the highest level in the processing hierarchy, information must pass not only through the retina, the thalamus and V1, but also V2, V4 and PIT. In addition, all these areas can function at the same time, thus allowing another type of parallelism, akin to the "pipeline" processing architectures that characterize many artificial vision systems.

The potential for using massively parallel processing has certainly been one of the main motivations for research in ANNs over the last 15 years. Indeed, there are now numerous pattern recognition systems based on neural networks and specific VLSI implementations have been developed.

2.2 The importance of rapid spike-based processing

However, there are numerous differences between current ANNs and processing in biological vision systems. One particularly interesting difference is that biological neurons use spikes to transmit information whereas the vast majority of ANNs are based on processing elements which either have binary outputs (0 or 1, as in the case of the original McCulloch-Pitts formulation) or continuous outputs, typically varying between 0 and 1. The use of a continuous variable to represent the output of a neuron has generally been thought to correspond to the firing rate of a biological neuron. Indeed, the

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 79-84

idea that neurons use firing rates to transmit information has been almost universally accepted since the first recordings from sensory nerves in the 1920s.

In the last few years, however, the universality of the "rate-coding" hypothesis has been increasingly called into question. In 1989 we had already argued that the speed with which the visual system could process information was effectively incompatible with the rate-coding hypothesis (Thorpe & Imbert, 1989). The basic argument can be summarized as follows. Neurons in the temporal lobe of the monkey can respond selectively to faces only 80-100 ms after stimulus onset (Oram & Perrett, 1992; Rolls & Tovee, 1994). To reach the temporal lobe in this time, information from the retina has to pass through roughly ten processing stages (see Fig. 1). Ten processing stages in around 100 ms leaves around 10 ms per processing layer, which, given that the firing rates of cortical neurons rarely exceed 100 spikes/s, makes it difficult to escape the conclusion that very few cells will be able to generate more than one action potential before the next stage has to respond. This clearly presents a serious problem if we wish to use firing rate as a code since one needs a minimum of two spikes in order to estimate the instantaneous firing rate of a cell.

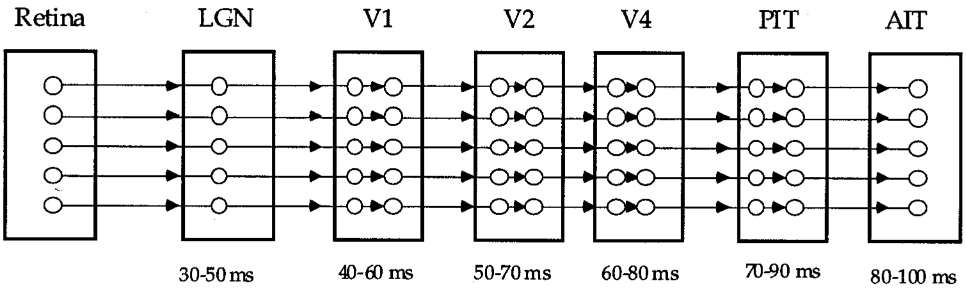

Figure 1 : Approximate latencies for neurons in different stages of the visual primate visual system (see Thorpe & Imbert, 1989; Nowak & Bullier, 1997). PIT - Posterior Inferotemporal cortex. AIT - Anterior Inferotemporal Cortex

The solidity of this temporal constraints argument has improved considerably since 1989. First, our recent ERP data has demonstrated convincingly that the rapid visual processing seen previously for faces also applies in the case of more general scene perception. Second, recent electrophysiological data has revealed that the conduction velocities of the fibres responsible for transmitting information between cortical areas may be much lower that previously thought - well under 1 m/s (Nowak & Bullier, 1997). Given that the shortest path from V1 to the temporal lobe in the monkey is of the order of 20-30 mm, it would appear that simple transmission will use up *at least* 20-30 ms of the time available for processing between V1 and IT. However, as illustrated in Figure 1, typical latency differences between V1 and IT are only about 40 ms, leaving very little time indeed for the actual processing to occur!

2.3 The importance of asynchrony

A second major difference between current ANNs and biological visual processing is that information transfer in biological systems is essentially *asynchronous.* We have just

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 79-84

seen how the speed of processing in the visual system is such that we are forced to conclude that only one spike may be available per neuron. It was in trying to come to terms with this constraint that we were led to propose an alternative way of coding information in neural networks based not on firing rates but rather on variations in the times at which different neurons fire (Thorpe, 1990, 1994). The basic idea is actually very simple. Traditionally, neurons as thought of as *analog-to-frequency* converters. However, we can also think of them as *analog-to -delay* converters since the time taken for an integrate-and-fire neuron to reach threshold in response to an input will depend on the strength of that input (see figure 2a), leading to the sort of non-linear intensity to latency transformation illustrated in figure 2b. Strong inputs will tend to produce short latency responses.

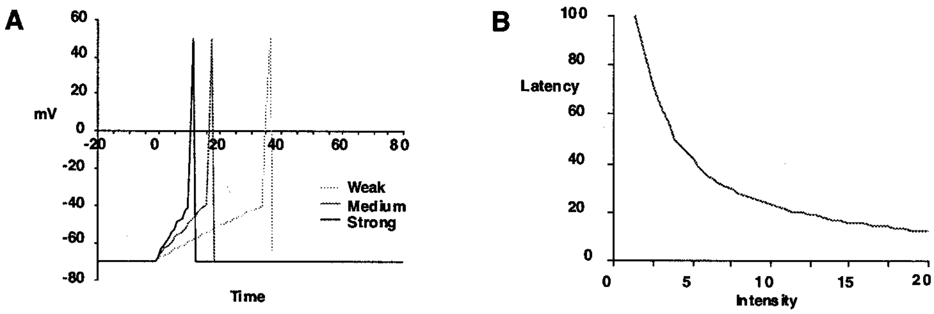

Figure 2. A . An integrate and fire neuron will reach threshold at different latencies depending on the strength of the input. **B .** A graph of latency as a function of input intensity for a typical integrate-and-fire neuron.

The idea that neurons can act as intensity-to-delay transcoders has important implications for understanding the nature of visual processing. First, it helps understand a puzzling feature of visual system neurophysiology, namely, the fact that the response latencies of neurons in any particular area are by no means fixed, but vary very considerably from neuron to neuron. In our studies of the responses of V1 neurons in the awake monkey to flashed gratings (Celebrini et al, 1993), we found that although some neurons could respond at around 40 ms after stimulus onset, most had latencies in the range 50-70 ms, and some had latencies of 100 ms or more. Similarly large latency ranges were seen using random dot stereograms as stimuli (Trotter et al, 1996). What could be the origin of this variability? While some of it is certainly due to factors such as the anatomical location of cell being recorded, and whether it is driven by the so-called magnocellular or parvocellular thalamocortical pathway (Nowak & Bullier, 1997), much of the variation could result from analog-to-delay transformations, in that neurons can only respond at short latencies when the stimulus is close to optimal. Neurons in V1 are sensitive to a large number of stimulus parameters, including not just orientation and position, but also spatial frequency, stereo disparity, colour and direction of movement. Most neurophysiological experiments can only hope to optimize one or two such parameters. Thus it is perfectly possible that a neuron that never responded with a latency shorter than say 80 ms in our experiments on orientation selectivity could have

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 79-84

responded much faster if, say, the colour of the stimulus has been optimized as well (all the stimuli in the Celebrini et al study were green!).

This raises the following intriguing possibility. Neurophysiological data indicate that only a very small percentage of cells in a particular cortical area will fire at short latencies in response to a particular stimulus. We saw earlier that rapid scene processing seems to depend essentially on only the first 5-10 ms of activity in any particular processing stage. However, there are good reasons to believe that the small number of cells that *do* fire in the first 5-10 ms are actually very special, because *they are the ones whose tuning is best matched to the input.* For this reason, the earliest processing in any particular area will be based on the responses of optimally activated cells in the previous layer.

Note that this notion of asynchronous processing also fits with another aspect of visual processing, namely the fact that object recognition appears to depend on the use of a relatively small subset of features. Anyone who has ever played Pictionary will know that most visual concepts can be defined using only a few lines - think of how a simple charicature can be enough to evoke the idea of, say, Bill Clinton. Thus, the major task for the visual system can be thought of as extracting just those visual features which are required for object recognition. It is here that the asynchronous propagation of information that characterizes the visual system will be so useful because it allows information to be "prioritized" in the temporal domain. In effect, we can suppose that the first information to reach any particular level in the processing hierarchy will be an extremely selective subset of the total information available. Further, the remarkable speed with which the visual system can function implies that, most of the time, the earliest arriving information allows us to make decisions that are nearly always correct!

3. SPIKENET - Towards asynchronous spiking networks

Over the last few years we have developed a novel type of neural network simulator with the aim of testing some of the ideas described in the previous section. SPIKENET is designed to allow the simulation of large numbers of integrate-and-fire neurons and in particular to explore the possibilities opened by the use of asynchronous spike-based propagation. Initial results have been very encouraging (Thorpe & Gautrais, 1997). So far we have developed simulations of the initial stages of visual processing, limited mainly to processing in the retina and visual cortex. However, we have been able to show that sophisticated visual processing is possible even under conditions where each neuron only has time to emit a single action potential. This therefore fits with the sorts of temporal constraints imposed by our experimental data of the speed of processing in the visual system. We have also shown that with the addition of further processing stages, it is possible to design neural network architectures capable of performing tasks such as digit recognition. Current work with Rufin VanRullen, a student in our group, is aimed at producing a neural network architecture capable of face recognition.

Finally, it is worth stressing that SPIKENET has numerous advantages from a purely computational point of view. The basic propagation process is in fact very simple - it consists essentially of maintaining a list of the neurons in the network which have just

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 79-84

spiked and propagating those spikes throughout the system. One consequence of this is that the process is computationally efficient. Even very large networks involving millions of units and hundreds of millions of connections can be simulated because the update process does not require recalculating the state of every unit as a function of all of their inputs (as with the vast majority of neural network simulators). Instead we simply propagate spikes through the network. This simplification is particularly important with multi-layer feed-forward nets because the later processing stages involve no significant computational overhead until the spikes have succeeded in traversing the earlier stages. Furthermore, spike-based simulators like SPIKENET are relatively easy to implement using parallel architectures because communication between processors can be restricted to sending lists of units which have fired. We are currently working on an implementation of SPIKENET based on PVM which will take advantage of this feature.

4. References

Celebrini S., Thorpe S., Trotter Y. & Imbert M. (1993). Dynamics of orientation coding in area V1 of the awake primate *Visual Neuroscience* **1 0**, 811-25.

Nowak L.G. & Bullier J (1997) The timing of information transfer in the visual system. In Kaas J., Rocklund K. & Peters A. (eds). Extrastriate Cortex in Primates (in press). Plenum Press.

Oram M. W. & Perrett D. I. (1992). Time course of neural responses discriminating different views of the face and head *Journal of Neurophysiology*, **6 8**, 70-84.

Rolls E. T. & Tovee M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking *Proc R Soc Lond B Biol Sci*, **2 5 7**, 9-15.

Thorpe S. J. (1990). Spike arrival times: A highly efficient coding scheme for neural networks. In R. Eckmiller, G. Hartman & G. Hauske (Eds.), *Parallel processing in neural systems* (pp. 91-94). North-Holland: Elsevier. Reprinted in H. Gutfreund & G. Toulouse (1994), *Biology and computation : A physicist's choice.* Singapour: World Scientific.

Thorpe, S., Fabre-Thorpe, M., & Richard, G. (1996). Rapid categorisation of natural images with extrafoveal presentations. *Perception*, **25** **Suppl.**, 45.

Thorpe S., Fize D. & Marlot C. (1996). Speed of processing in the human visual system *Nature*, **3 8 1**, 520-522.

Thorpe, S. J., & Gautrais, J. (1997). Rapid visual processing using spike asynchrony. In M. Jordan (Ed.), *Neural Information Processing Systems* (Vol. 9, in press): M.I.T. Press.

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

Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S., & Imbert, M. (1996). Neural processing of stereopsis as a function of viewing distance in primate visual cortical area V1. *Journal of Neurophysiology* **7 6**, 2872-2885