

Face identification using one spike per neuron: resistance to image degradations

Arnaud Delorme* and Simon J. Thorpe

Centre de Recherche Cerveau & Cognition UMR 5549, 133, route de Narbonne, 31062 Toulouse, France
arno@cerco.ups-tlse.fr, thorpe@cerco.ups-tlse.fr

ABSTRACT The short response latencies of face selective neurons in the inferotemporal cortex impose major constraints on models of visual processing. It appears that visual information must essentially propagate in a feed-forward fashion with most neurons only having time to fire one spike. We hypothesize that flashed stimuli can be encoded by the order of firing of ganglion cells in the retina and propose a neuronal mechanism, that could be related to fast shunting inhibition, to decode such information. Based on these assumptions, we built a three-layered neural network of retino-topically organized neuronal maps. We showed, by using a learning rule involving spike timing dependant plasticity, that neuronal maps in the output layer can be trained to recognize natural photographs of faces. Not only was the model able to generalize to novel views of the same faces, it was also remarkably resistant to image noise and reductions in contrast.

Electrophysiological studies indicate that neurons in the inferotemporal cortex (IT) respond selectively to faces only 80-100 ms after stimulus presentation (Bruce, Desimone, & Gross, 1981; Perrett, Rolls, & Caan, 1982; Jeffreys, 1996). Within this time, visual information has to travel through many hierarchical layers in the visual system: from the retina, spikes propagate to the lateral geniculate nucleus (LGN), then to cortical visual areas V1, V2 and V4 before reaching higher visual areas in the anterior and posterior inferotemporal cortex (respectively AIT and PIT). With at least 2 synaptic stages per cortical stage and a mean synaptic conduction and integration time of about 10 ms (Nowak & Bullier, 1997), such data imply that the neurons at most processing stages will only rarely be able to fire more than one spike before the next stage has to respond (Thorpe & Imbert, 1989). Given that at least two spikes would be necessary to estimate spike frequencies, this raises severe problems for the conventional view that the neurons are transmitting the information in the form of a rate code. It has been shown that, after a flash, the first wave of spikes can carry a lot of information about the stimulus (Heller, Hertz, Kjaer, & Richmond, 1995; Tovee & Rolls, 1995; Sugase, Yamane, Ueno, & Kawano, 1999). In V1, stimuli, presented for 10 ms and followed by a mask, can still trigger responses that are orientation selective (Celebrini, Thorpe, Trotter, & Imbert, 1993). In IT, under similar circumstances (14 ms between images), neurons can still respond selectively to faces (Keysers, Xiao, Foldiak, & Perrett, 2001). Moreover, psychophysical evidence also suggests that rapid visual categorization depends mainly on feed-forward processing (Thorpe, Fize, & Marlot, 1996; Delorme, Richard & Fabre-Thorpe, 2000) and is no faster for highly familiar images than for ones that have never been seen before (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001). Thus,

biological and psychophysical studies seem to agree that highly selective responses in the visual system can be produced using essentially automatic feed-forward processing.

We thus need to find neuronal codes consistent with such constraints. We have argued elsewhere that the use of relative latency coding, in which the order of firing across a population of neurons is used to encode flashed stimuli, offers many advantages (Thorpe, 1990; Gautrais & Thorpe, 1998, Van Rullen & Thorpe, 2001). It is compatible with the constraint of using only one spike at each processing stage and seems very powerful in terms of information encoding: a population of N neurons can actually discriminate $N!$ stimuli whereas, within the same time window, a more classical population rate-code approach could only encode $N+1$ stimuli (Gautrais & Thorpe, 1998).

In order to investigate the power of this form of coding, we ran simulations using SpikeNet (Delorme, Gautrais, VanRullen, & Thorpe, 1999), a software package designed for modeling networks containing hundreds of thousands of asynchronously firing integrate-and-fire units. We have already shown that such networks are able to detect faces in natural photographs (VanRullen, Gautrais, Delorme, & Thorpe, 1998). In this present paper, we go further by demonstrating the ability of SpikeNet based networks to perform a much more challenging face identification task. The network was required to determine the identity of a person from novel views that were not presented during learning. We also analyzed the performance of the network with noisy and low contrast inputs.

1. Architecture of the model

With this model, our goal is to demonstrate, within a neurobiologically plausible framework, the ability of a

network that uses only one spike per neuron to process faces (and quite probably other classes of stimuli) in natural photographs. The network is hierarchically organized into three layers of retinotopic maps containing relatively simple integrate-and-fire neurons. The model was kept as simple as possible but is roughly based on the architecture of the primate visual system, with a first layer corresponding to the retina, the second one for V1 and the last one for V4-IT. The pattern of connectivity becomes increasingly complex as processing reaches higher levels. Spikes were propagated in a feed-forward manner through the whole network. A major constraint was that, at all levels of processing, from the retina to higher neuronal maps, neurons cannot spike more than once, thus preventing the use of conventional rate-based coding schemes. Moreover, because iterative loops cannot occur, the propagation dynamics were purely feed-forward.

The retina layer included ON and OFF center cells whose activation levels depended on the local contrast at a given location in the input image (difference of gaussian 3x3, normalized to 0). At each location in the input image, there was a pair of ON and OFF-center ganglion cells only one of which was allowed to fire. The neurons used a simple integrate-and-fire mechanism that meant that spike latency is inversely proportional to the activation value. This means that the earliest cells to fire will correspond to the parts of the input image where the contrast is the highest (figure 1). We will see later that, as long as the relation

between latency and contrast is monotonously decreasing, the exact transformation function does not alter the propagation.

In the second layer of the model, neurons had orientation selectivity (8 different orientations separated by 45°). The pattern of connectivity was implemented using Gabor functions ($\sigma=1$ neuronal unit; $\phi = 0.5$ rad/neuronal unit) similar to those used in previous studies (VanRullen et al., 1998). The neuronal thresholds were all equal and adjusted in such a way that for a given image input, only about 10-20% of the neurons produced a spike.

In the last layer, the number of neuronal maps corresponded to the number of individuals presented to the network. Neurons were trained to respond selectively to the presence of a given person at the center of their receptive field (which include most of the input image) and whenever a neuron spiked, it inhibited all the neurons of the other neuronal maps in a zone centered on the neuron's location (Gaussian distribution of synaptic weights of width $\sigma=2$ units). The inhibition was strong enough to prevent units in other maps from firing. Thus, neuronal discharges can be seen to be selective to the presence of a given person at one location in the input image.

2. Neurons

Neurons were simple integrate-and-fire units: they integrated afferent spikes until they reached a threshold and fired once. The latency of discharge of the output

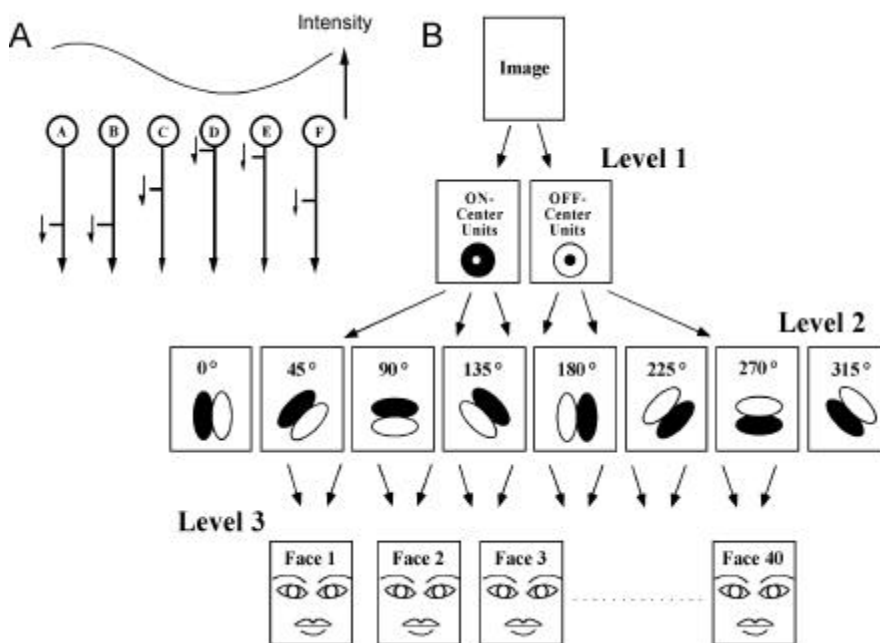


Figure 1: A. Neurons can be considered as integrate-and-fire units with spike latencies that depend on the local activation. Strongly activated neurons will fire first ($B > A > F > C > E > D$) and with only 6 neurons, one can encode $6!$ (i.e. 720) stimulus profiles. B. Architecture of the model. It was built of three processing stages: the image was first decomposed using ON-center and OFF-center contrast filters whose outputs are used to determine spike latencies. In a model of V1, 8 maps of orientation selective cells (each separated by 45°) integrated these spikes. In the last layer, which corresponds to V4-IT, neurons were selective to faces (one neuronal map for each individual). Spike propagation is feed-forward only and iterative processes can not occur in the sense that, even if lateral interactions are present in the last processing stage, each neuron can only fire once.

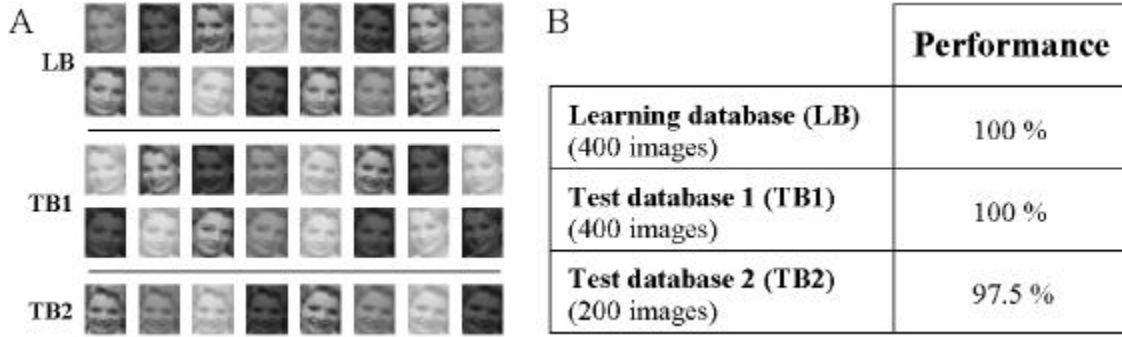


Figure 2: A. Illustration of the database of images used in the learning procedure (LB) and for the first and second test databases (TB1 and TB2). Database TB1 contained the same views as in the learning set but had different contrast and luminance values. The second test base (TB2) contained novel views of each individual again with varying contrast and luminance values. B. Results of the model on these databases. The performance on the LB and TB1 was perfect. On TB2, the drop in performance was only 2.

neuron depended upon the relative order of firing of its afferent in the following way: let $A = \{ a_1, a_2, a_3 \dots a_{m-1}, a_m \}$ be the ensemble of afferent neurons of neuron i and $W = \{ w_{1,i}, w_{2,i}, w_{3,i} \dots w_{m-1,i}, w_{m,i} \}$ the weights of the m corresponding connections; let $\text{mod} \in]0, 1[$ be an arbitrary modulation factor. Each time the neuron receives a spike, the efficiency of spike integration is divided by this factor, with the result that the earliest spikes have the strongest impact on the post-synaptic potential. Such a mechanism implements a general decoding scheme for input latencies (Thorpe and Gautrais, 1998). In the current simulations, the modulation factor was set so that when half the active inputs have fired, the effectiveness of any particular input is reduced by 50%. The activation level of neuron i at time t is given by

$$\text{Activation}(i, t) = \sum_{j \in [1, m]} \text{mod}^{\text{order}(a_j)} w_{j,i}$$

where $\text{order}(a_j)$ is the firing rank of neuron a_j in the ensemble A. By convention, $\text{order}(a_j) = +8$ if neuron a_j has not fire at time t , setting the corresponding term in the above sum to zero. This kind of desensitization function could correspond to a fast shunting inhibition mechanism.

Neuron i will fire at time t if (and only if)

$$\text{Activation}(i, t) \geq \text{Threshold}(i)$$

Under such conditions, two key features can be pointed out. First, the activation of the neuron is highest when the order of afferent discharges matches the pattern of weights. If the highest synaptic weights are activated first, their effectiveness was unaffected by the modulation. Second, because of this kind of spike integration, the most strongly activated neurons fired first. These points have important consequences because the performance of the model depends on the dynamics of this process.

3. Learning procedure

The image database included 400 faces resized to 28x23 pixels (10 views of 40 persons) corresponding to the whole AT&T Cambridge Laboratories face database (formerly “the ORL face database” available at <http://www.uk.research.att.com/facedatabase.html>). Individuals were of both sexes, from different origins, with or without various characteristics such as glasses, beard or moustache. Views were frontal ($\pm 30^\circ$).

We constructed 3 databases of images (figure 2). Out of the 10 views of each given individual, 8 were randomly selected to build the image database used for the learning phase and the first test phase. To test the robustness of the model to contrast and luminance changes, 3 additional versions of all 400 views were generated. One set had half contrast (pixel values were recalculated to be in the range 64-191 over mean gray level 128). The other two also had half contrast but the gray levels were shifted to either higher or lower luminance levels by adding or subtracting 64. Among these 4 versions, 2 were randomly chosen to be used during learning and the remaining 2 were used to build a first test base (thus there were 8x2 images of each individual in both cases).

Of the 10 views of each individual, 8 had already been used for learning with the first test base. The remaining 2, together with their 3 additional versions, were used to test the model with views of each individual that had never been presented (2x4 images for each of the 40 individuals).

Learning was supervised and implemented as follows: first, before the propagation of each image of the learning database, the neuron in the last layer that corresponds to the location of the center of the face presented was preselected. The center of the face was defined as the isobaric center of the nose and the two eyes. These locations were determined by hand by clicking on these features of the image at high resolution. During the propagation of an image, the synaptic weight distribution of inputs to the selected

neuron was modified according to the discharge order of the afferents.

More specifically, for a synaptic weight between neuron j and the preselected neuron i :

$$\Delta w_{j,i} = \frac{\text{mod}^{\text{order}(a_j)}}{N}$$

with the same convention as previously, N being the number of images of each individual (10 in our case). As we pointed out previously, after the learning phase, the neuron is more selective to the pattern that was presented because the highest synaptic weights tend to

corresponded to the afferent neurons that fired first. Moreover, because the neurons in the output maps share the same set of synaptic weights, responses of neuronal maps were invariant to the location of the face in the image. Whenever a neuron's synaptic weight was modified, it affected all the neurons of the map and each weight converged to a value that depended on the mean rank of each input to the neuronal map (in the network, the synaptic weight value at the end of the learning phase was proportional to the mean modulation of the synapse). Thus, within a face selective neuronal map, all the neurons in the output maps became selective to the "average" view of one individual (figure 3).

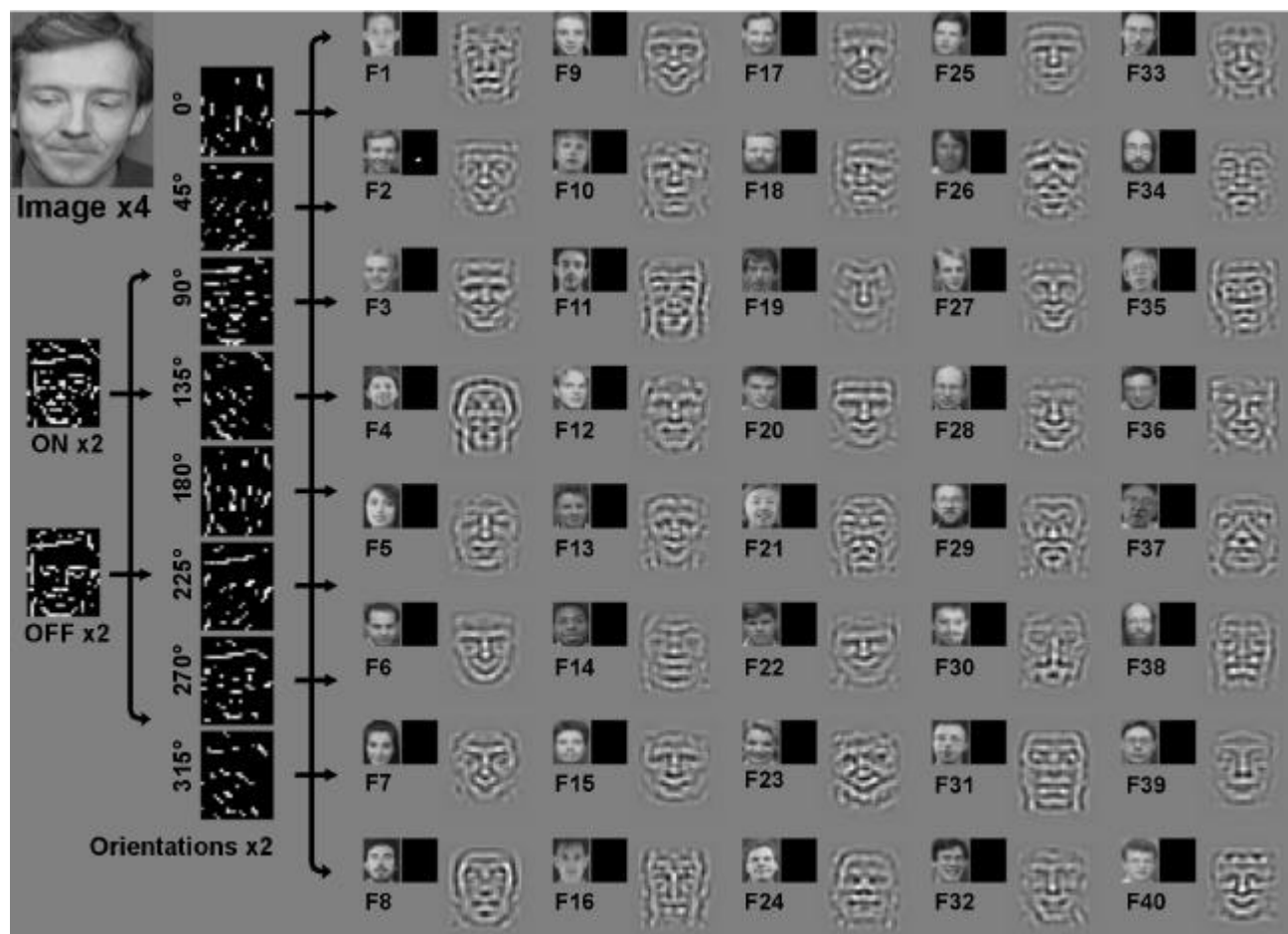


Figure 3: Illustration of the propagation of a single face in the network. In the top left corner, an image is presented to the network (x4 indicating that the actual size of the image was 4 times smaller). It is first decomposed into ON and OFF center contrast and then into orientation filters (magnified twice for better legibility). The light pixels indicate neurons that were activated at short latencies, whereas darker pixels correspond to neurons that fired progressively later during propagation. F1 to F40 display the spike activity for the 40 neuronal maps selective to the 40 individuals. The only neurons to discharge were those at the center of the neuronal map selective to the individual in the input image (F2). For each neuronal map, a reconstruction of the selectivity of its neuron is also indicated on the right of its discharge map (magnified twice for better readability). For a neuron of one face selective map, this reconstruction corresponds to a linear combination of input neurons' selectivity (i.e. Gabor patches at various orientation and positions relative to the output neuron location) weighted by the synaptic strengths connecting these input neurons to the output neuron.

4. Thresholds

Because of the large number of neuronal maps selective to faces, the thresholds of output neurons cannot easily be adjusted by hand as was the case in the previous studies (VanRullen et al., 1998). Instead, we used an optimization procedure that adjusted the threshold of each target map individually so that each neuronal map responded to the same proportion of images in the database, regardless of whether they contained the individual to which they were selective (with 40 neuronal maps, this corresponded to a probability of 2.5%). A map was considered as responding to a given image when the first neuron to fire in the last layer belonged to this map. We expected that, if the learning phase was accurate, this procedure would result in neuronal maps that were selective to

images that contained the views of persons they were trained on. This algorithm offers the great advantage of not being supervised.

5. Results

The accuracy of the network was determined by computing the rate of correct responses relatively to the number of images (thus random responses would lead to 2.5% correct responses). As in the case of threshold optimization, we considered a neuronal map to respond to a specific image if it contained the first face selective neuron that discharged. The pattern of results on the three databases is presented in figure 2. It shows that the recognition accuracy on the database used for learning was 100% correct. Performance was also 100% correct when using the first testing database



Figure 4: Propagation of the whole set of 400 initial images (belonging either to the learning base or one of the two test bases). The network is a scaled version of the one in figure 3. Here, we only presented the global result of the network superposed with the input image. A ray-tracing algorithm was used to fill the spaces between each face image in the montage. The size of the image was 910x700 pixels, which requires a network containing roughly 32 million neurons and 245 billion connections. For a correct detection, a neuron selective to a particular face must discharge within a 4x4 region located at the center of the face. Black squares indicate correct recognition and white ones false detections. Despite the size of the network, the simulation could be completed in about 30 minutes of CPU time on a modest desktop computer (Macintosh G3, 266 MHz).

that contained views from the learning set but with different contrast and luminance, thus demonstrating that performance was robust to contrast and luminance modifications. Using the second test database, the percentage of correct responses reached 97.5%. This database composed of views of individual that had not been presented during the learning phase and thus reflects the ability to the network to generalize to new views.

In all the previous simulations, each 23 by 28 pixel input image contained only one face. Since the total network size was $(2 + 8 + 40)$ times the number of pixels in the image, the total number of neurons was 32200 neurons. However, by simply changing the size of the input image, and thus scaling up the network, we were able to test the network with a very large image that contained all the 400 original views of the faces and involved roughly 32 million neurons. As illustrated in Figure 4, the results show that even under these conditions, the model was able to simultaneously process all the faces and maintain the accuracy of identification at over 98%.

6. Resistance to noise and contrast reduction

To test the robustness of the model, the images were degraded by lowering contrast or adding noise. The decrease in contrast was achieved by limiting the range of pixel values in the images of the learning database. The results showed that the network performed well even with substantial reductions in image contrast. Indeed, only when contrast was reduced to below 3%

did the network fail (figure 5). At that level of contrast, the gray levels in the image were restricted to only 1-5 possible values. Performance was also studied when noise was added to the images during learning (weighted average of the initial image with an image made of pixels that took random gray values). As indicated in figure 5, even with noise, the performance of the network was still impressive: with 50% noise, performance is still above 80% correct.

It is worth noting that these high levels of performance were obtained despite the fact that the discharge probability was not adjusted for each condition of noise or contrast. The thresholds of neurons in the face selective maps were fixed, based on the discharge probability on the learning database, and were kept constant for the rest of the simulation. For instance, with a high percentage of noise, orientation selective neurons in the second layer did not fire any more, so face selective neuronal maps could not integrate any inputs. Similarly, with 1% of residual contrast, there was simply no further activity in the face-selective output maps. Lowering the threshold of these output maps may well allow performance to be improved even further.

7. Biological relevance

We have already reviewed some of the arguments in favor of the feed-forward propagation that was implemented in our model (Celebrini et al., 1993; Thorpe et al., 1996; Delorme et al., 2000; Fabre-Thorpe et al., 2001). Special emphasis should be given

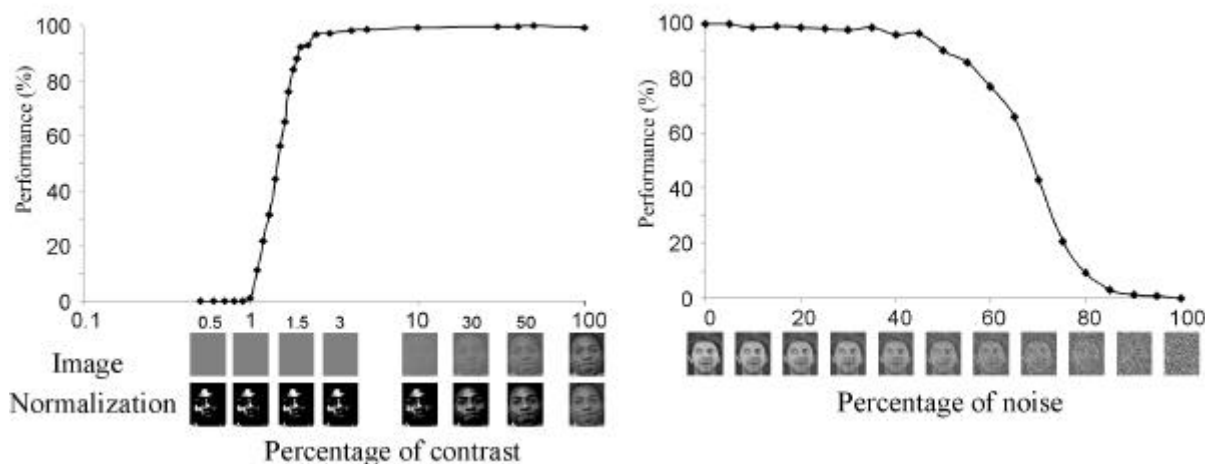


Figure 5: Resistance to noise and contrast reduction. A. Contrasts were progressively reduced for the images in the learning database (the contrast reduction corresponded to a reduction in the range of pixel values around a mean gray level 128). This curve shows that accuracy was still above 95% even when contrast was reduced to 2%. The faces below the abscissa illustrate the contrast reduction and normalized versions of the images indicate the remaining information available in the image. B. Noise was introduced in the images of the learning database by a weighted average of the initial image with an image made of pixels that took random gray values. Thus with 100% noise, no more information about the initial image was present. This curve shows that even with a noise level of 45%, the accuracy of the network was still above 96%. The images below the abscissa illustrate the deterioration with noise. For both contrast and noise condition, the images used to illustrate the effects were those which induced the most resistant responses.

to recent studies that show that, when presented with a face to which it responded selectively, neurons in inferotemporal cortex remained selective even when images were successively presented in a RSVP (Rapid Serial Visual Presentation) sequence during only 14 ms each (Keyser et al., 2001). Under these constraints, when a neuron responded to a face presented 80-100 ms earlier, V1, V2 and V4 were presumably still processing the 6-7 subsequent images. Such data argue strongly in favor of a mainly feed-forward processing strategy. The simulations presented here demonstrate that a simple processing mechanism compatible with these temporal constraints is nevertheless capable of producing neural responses that are surprisingly selective. Clearly, we would not wish to argue that the present simulations provide a realistic view of how face selective responses are produced. On the other hand, they do suggest that the computational capacities of simple feed-forward networks of asynchronously firing neurons have been seriously underestimated in the past.

Although there is evidence that single spikes can be reliable and carry considerable amounts of information (Mainen & Sejnowski, 1995; Buracas, Zador, DeWeese, & Albright, 1998), it is still generally believed that most of the useful information is encoded in the rate of discharge or in bursts. We would like to argue that, at least for bursts, this is not contradictory with our hypothesis. Bursts can be considered as single events with the exact time and number of spikes within a burst being rather uninformative. A number of studies suggest that the latency of the first spikes that carries considerable information about the stimulus (Livingstone, Freeman, & Hubel, 1996; Berry, Warland, & Meister, 1997; Lisman, 1997; Reinagel, Godwin, Sherman, & Koch, 1999). Because of the fast depression of excitatory synapses (Markram & Tsodyks, 1996), the approximation of a single spike event for a burst may be sufficient and would thus correspond to the spike integration mechanism implemented in our network.

More specifically, we need to point out some of the major simplifications used in the simulations. A first point is that we did not include a leakage term – the neurons simply summed the incoming spikes, modulating the effectiveness of each input by a factor that depended on the order of that input. While this is clearly unrealistic, it seems likely that when processing is very rapid, the effect of including a leakage term based on a typical membrane time constant of say 10-20 ms would be minor. Furthermore, by avoiding leakage currents, the responses of the neurons become virtually invariant with changes in contrast, since the final activation state of the output neurons depends only on the order in which the inputs fire, and not their precise latency. The use of this simulation has the added benefit that the precise form of the intensity-latency function is much less critical since any

transformation of contrast to latency that corresponds to a monotonous decreasing function would have given the same result. For similar reasons, any desensitization function that is a monotonic decreasing function of the number of spikes would give effectively the same pattern of results (Gautrais & Thorpe, 1998; VanRullen & Thorpe, 2001). Note also, that the fast desensitization mechanism we used fits well with intracellular recording studies that show that after a flash, neuronal conductance changes occur very rapidly in a few ms and could be related to stimulus-driven shunting inhibition (Borg-Graham, Monier, & Fregnac, 1998). Moreover, it is now accepted, at least in V1, that, after a flash, the integration of excitatory post-synaptic potentials (EPSPs) interacts with fast inhibition (Gabbott, Martin, & Whitteridge, 1988; Celebrini et al., 1993; Hirsch, Volgushev, Pei, Vidyasagar, & Creutzfeldt, 1993; Alonso, Reid, & Martinez, 1998) even for the first spike (Hirsch et al., 1998).

The architecture used in the current simulations was, with the exception of the inhibitory connections between retinotopically corresponding zones of the output maps, a pure feed-forward one. Note, however, that processing based on a wave of spike propagation in which no individual neuron fires multiple spikes could also include contributions from lateral connections (VanRullen, Delorme, & Thorpe, 2001). We would argue that, as long as no individual cell is required to emit more than one action potential, the network should be considered as functionally feed-forward, even though anatomically defined feedback loops are present. This argument might even apply to connections between processing layers, for instance between V2 and V1. Rapidly occurring effects of anatomical feedback (Hupé, James, Payne, Lomber, Girard, & Bullier, 1998) could still fit within the processing wave model proposed here as long as the effects of the top-down activity occur before some V1 neurons have emitted their first spike. For this reason, the sort of processing mode proposed here is not incompatible with the large amount of anatomical feedback and lateral connectivity in the visual system. Nevertheless, there are undoubtedly other important roles for top-down connections that would include attentional modulation and possibly learning.

The learning rule we used, optimal in terms of the rank order coding hypothesis, could be seen as the most artificial part of our network. However, we showed recently that, because the speed of the integration process is fast compared to the dynamics of reinforcement/depression, our learning rule could be linked to spike timing neuronal plasticity (Delorme et al., 2001). This kind of neuronal plasticity indicates that the order of input and output spikes is critical in determining the weight change: if the EPSP occurs before the postsynaptic neuron spikes, the synapse is strengthened, otherwise it is depressed (Markram,

Lubke, Frotscher, & Sakmann, 1997; Bi & Poo, 1998). Such a rule might well fit with the idea of increasing the weights of inputs that are systematically among the first to fire. Finally, although part of the learning process was supervised, the optimization procedure was very simple, and served only to equalize discharge probability in the output maps. No attempt was made to adjust relative thresholds specifically with respect to accuracy.

8. Performance of the model

Responses of our network seem remarkably accurate when compared to other models (for a review see Gong, McKenna, & Psarrou, 2000). Face recognition models often use complex dynamic link matching mechanisms where the image is transformed to obtain a version which is more or less invariant to the view presented (Wiskott & von der Malsburg, 1995; Würtz, 1997). Here we showed that this type of preprocessing might not be necessary, since the model is able to achieve limited view invariance without any of these mechanisms. Thus, our model is more related to standard principal component analysis (Fogelman-Soulie, Viennet, & Lamy, 1993; Valentin, Abdi, O'Toole, & Cottrell, 1994) or statistical networks (Samaria & Harter, 1994). However, the performance we obtained appears to be quite different from these models. For instance, on the same image database we used, a model based on Hidden Markov Chains only reached an accuracy level of 90% on novel views of the person presented during learning (Samaria & Harter, 1994). Although the exact image distribution in the learning database differed from our simulation, the network presented here achieved 97.5% correct under similar conditions. Another network by Mel (1997), composed of two layers, also performs well in the categorization of novel views of objects. The first layer extracts basic features (orientation, color, blobs, vertices...) and the second one implements a powerful classifier based on a nearest neighbor classifier technique derived from artificial intelligence. However, although the performance of this network on the classification of novel views was impressive (97% of correct responses), it dropped very rapidly when images were altered (80% of correct responses if the color was removed; 58% under 30% of noise). Although direct comparisons are always difficult, the performance of our network appears more robust since it does not rely on color cues and still reach 98% of correct responses under 30% of noise.

Our approach, though very simplified, shows that one can achieve high performance on challenging image processing tasks using a simple feedforward architecture based on biological vision. It also suggests that computational neuroscience would greatly benefit from paying close attention to the computational advantages associated with spike-based processing. This conclusion should not be surprising for the

biologist, since millions of years of evolution have optimized the visual system to achieve the best compromise between accuracy and processing speed.

This report was supported by a grant from INRIA in France (Institut National de Recherche en Informatique et Automatique).

- Berry, M. J., Warland, D. K., & Meister, M. (1997). The structure and precision of retinal spike trains. *Proceedings of the National Academy of Science U S A*, 94 (10), 5411-5416.
- Bi, G. Q., & Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18 (24), 10464-10472.
- Borg-Graham, L. J., Monier, C., & Fregnac, Y. (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, 393 (6683), 369-373.
- Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, 46 (2), 369-384.
- Buracas, G. T., Zador, A. M., DeWeese, M. R., & Albright, T. D. (1998). Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron*, 20 (5), 959-969.
- Celebrini, S., Thorpe, S. J., Trotter, Y., & Imbert, M. (1993). Dynamics of orientation coding in area V1 of the awake primate. *Visual Neuroscience*, 10 (5), 811-825.
- Delorme, A., Gautrais, J., VanRullen, R., & Thorpe, S. J. (1999). SpikeNET: A simulator for modeling large networks of integrate-and-fire neurons. *Neurocomputing*, 26-27, 989-996.
- Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Res*, 40(16), 2187-2200.
- Delorme, A., Perrinet, L., & Thorpe, S. J. (2001). Network of integrate-and-fire neurons using rank order coding I: Receptive field emergence. *Neurocomputing*, in press.
- Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. J. (2001). A limit to the speed of processing in Ultra-Rapid Visual Categorization of novel natural scenes. *Journal of Cognitive Neuroscience*, in press.
- Fogelman-Soulie, F., Viennet, E., & Lamy, B. (1993). Multi-modular neural network architectures: applications in optical character and human face recognition. *International Journal of Pattern Recognition and Artificial Intelligence*, 8, 147-165.
- Gabbott, P. L., Martin, K. A., & Whitteridge, D. (1988). Evidence for the connections between a clutch cell and a corticotectal neuron in area 17 of the cat visual cortex. *Proceedings of the Royal Society of London (B)*, 233 (1273), 385-391.
- Gautrais, J., & Thorpe, S. (1998). Rate coding versus temporal order coding: a theoretical approach. *Biosystems*, 48 (1-3), 57-65.
- Gong, S., McKenna, S. J., & Psarrou, A. (2000). *Dynamic Vision: From Images to Face Recognition*. London: Imperial College Press.
- Heller, J., Hertz, J. A., Kjaer, T. W., & Richmond, B. J. (1995). Information flow and temporal coding in primate pattern vision. *Journal of Computational Neuroscience*, 2 (3), 175-193.
- Hirsch, J. A., Alonso, J. M., Reid, R. C., & Martinez, L. M. (1998). Synaptic integration in striate cortical simple cells. *Journal of Neuroscience*, 18 (22), 9517-9528.
- Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, 394(6695), 784-787.
- Jeffreys, D. A. (1996). Evoked potential studies of face and object processing. *Visual Cognition*, 3, 1-38.
- Keysers, C., Xiao, D., Foldiak, P., & Perrett, D. I. (2001). The Speed of Sight. *Journal of Cognitive Neuroscience*, in press.

- Lisman, J. E. (1997). Bursts as a unit of neural information: making unreliable synapses reliable. *Trends in Neuroscience*, 20 (1), 38-43.
- Livingstone, M. S., Freeman, D. C., & Hubel, D. H. (1996). Visual responses in V1 of freely viewing monkeys. *Cold Spring Harbor Symposium of Quantitative Biology*, 61, 27-37.
- Mainen, Z. F., & Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, 268 (5216), 1503-1506.
- Markram, H., Lubke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275 (5297), 213-215.
- Markram, H., & Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, 382 (6594), 807-810.
- Mel, B. W. (1997). SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Computation*, 9 (4), 777-804.
- Nowak, L. G., & Bullier, J. (1997). The timing of information transfer in the visual system. In J. Kaas, K. Rocklund, & A. Peters (Eds.), *Extrastriate cortex in primates* (pp. 205-241). New-York: Plenum Press.
- Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, 47 (3), 329-342.
- Reinagel, P., Godwin, D., Sherman, S. M., & Koch, C. (1999). Encoding of visual information by LGN bursts. *Journal of Neurophysiology*, 81 (5), 2558-2569.
- Samaria, F. S., & Harter, A. C. (1994). Parametrisation of a Stochastic Model for Human Face Identification. *Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision*, Sarasota.
- Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, 400 (6747), 869-873.
- 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). Amsterdam: Elsevier.
- Thorpe, S. J., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381 (6582), 520-522.
- Thorpe, S.J. and Gautrais, J. (1998). Rank Order Coding. In J. Bower, (Ed.) *Computational Neuroscience: Trends in Research 1998*, (pp. 113-118.) New York, Plenum Press.
- Thorpe, S. J., & Imbert, M. (1989). Biological constraints on connectionist models. In R. R. Pfeifer, Z. Schreter, F. Fogelman-Soulié, & L. Steels (Eds.), *Connectionism in Perspective* (pp. 63-92). Amsterdam: Elsevier.
- Tovee, M. J., & Rolls, E. T. (1995). Information encoding in short firing rate epochs by single neurones in the primate temporal cortex. *Visual Cognition*, 2, 35-59.
- Valentin, D., Abdi, H., O'Toole, A., & Cottrell, G. W. (1994). Connectionist models of face processing: a survey. *Pattern Recognition*, 27, 1209-1230.
- VanRullen, R., Delorme, A., & Thorpe, S. J. (2001) Feed-forward contour integration in primary visual cortex based on asynchronous spike propagation. *Neurocomputing*, in press.
- 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. J. (2001). Rate coding versus temporal order coding: What the retinal ganglion cells tell the visual cortex. *Neural Computation*, 13(6) in press.
- Volgushev, M., Pei, X., Vidyasagar, T. R., & Creutzfeldt, O. D. (1993). Excitation and inhibition in orientation selectivity of cat visual cortex neurons revealed by whole-cell recordings in vivo. *Visual Neuroscience*, 10 (6), 1151-1155.
- Wiskott, L., & von der Malsburg, C. (1995). Recognizing faces by dynamic link matching. *Proceedings of the ICANN '95*, Paris (pp. 347-352).
- Würtz, R. P. (1997). Neuronal theories and technical systems for face recognition. *Proceedings of the ESANN*, Brussels (pp. 73-79).