Stimulus detection after interruption of the feedforward response in a backward masking paradigm

August Romeo · Maria Sole Puig · Laura Pérez Zapata ·
Joan Lopez-Moliner · Hans Supèr

Received: 28 June 2011 / Revised: 11 January 2012 / Accepted: 31 January 2012 / Published online: 14 February 2012
© Springer Science+Business Media B.V. 2012

Abstract In backward masking, a target stimulus is rendered invisible by the presentation of a second stimulus, the mask. When the mask is effective, neural responses to the target are suppressed. Nevertheless, weak target responses sometimes may produce a behavioural response. It remains unclear whether the reduced target response is a purely feedforward response or that it includes recurrent activity. Using a feedforward neural network of biological plausible spiking neurons, we tested whether a transient spike burst is sufficient for face categorization. After training the network, the system achieved face/non-face categorization for sets of grayscale images. In a backward masking paradigm, the transient burst response was cut off thereby reducing the feedforward target response. Despite the suppressed feedforward responses stimulus classification remained robust. Thus according to our model data stimulus detection is possible with purely, suppressed feedforward responses.

Keywords Computational modeling · Backward masking · Face detection · Face categorization

Introduction

To know whether feedforward activity is capable of detecting sensory information, backward masking experiments are used that intend to separate feedforward evoked stimulus responses from feedback modulatory effects. However, stimulus-evoked responses e.g. of face selective cells outlast the presentation period (Quiñero et al. 2008). So despite the short stimulus presentation times in masking, neurons may engage in recurrent processing.

Computer models show that a single spike is sufficient for object classification (VanRullen et al. 1998; Delorme and Thorpe 2001). However, using a simple spiking model these studies contrast objects with clearly different non-object images and focus on temporal coding where the earliest stimulus evoked spikes have a strong influence on detection performance. Therefore, it is arguable that the model is analogue to artificial neural networks. Furthermore, these studies employ complex receptive field features that are designed and strategically located to detect face components. Such detailed information appears to be encoded in the late stimuli evoked responses, i.e. subject to recurrent processing, of IT neurons (Sugase et al. 1999). It is therefore not clear whether these models resemble purely feedforward processing in the brain or implicitly include recurrent processing.

We designed a computational feedforward model using biologically plausible spiking neurons where all spikes have an equal effect on the classifier cell and tested it for face categorization after backward masking. After training the network, the model achieved face/non-face categorization for sets of face images. After probing the model in a backward masking paradigm we observed that the spike burst to the face/non-face image was disrupted. This disruption was caused by the removal of the stimulus and not

A. Romeo · M. S. Puig · L. P. Zapata · J. Lopez-Moliner ·
H. Supèr (✉)
Department of Basic Psychology, Faculty of Psychology,
University of Barcelona (UB), Pg. Vall d' Hebron 171,
08035 Barcelona, Spain
e-mail: hans.super@icrea.es

M. S. Puig · L. P. Zapata · J. Lopez-Moliner · H. Supèr
Institute for Brain, Cognition and Behavior (IR3C),
Barcelona, Spain

H. Supèr
Catalan Institution for Research and Advanced Studies (ICREA),
Barcelona, Spain

by the presentation of the mask. Despite the short period for integrating spiking input, face detection remained robust. We complemented the model findings with a human psychophysical backward masking test using the same face/non-face images. Our model observations indicate that classification is possible with strongly suppressed feed-forward responses.

Methods

Face database

In the present work we employed sets of grayscale images (19×19 pixels), with frontal and near-frontal views of faces, from the CBCL database¹ (Weyrauch et al. 2004). Images were drawn randomly from the CBCL database and their pixel values were rescaled to the standardized range $[-1, +1]$.¹

Model

The input layer is an array “X” containing the rescaled values of the image, which feeds into an intermediate layer, L0, which converts the image input into a spike map (Fig. 1). There are as many weights from the input layer to the layer L0 as pixels in the image. All the connections (weights) from the stimulus to layer L0 have the same weight but the input is variable as it is a grey scale image. The output layer consists of a single unit or ‘output neuron’ (ON). Our two layer model can in principle be applicable to any level of visual processing, like V1 or IT. Extra layers can be added to the model without changing the results. The extra layers just consistently transmit the spike map to the next higher level, e.g. IT.

Neural dynamics of the spiking model

We used the spiking neurons of Izhikevich (2003). In previous work we demonstrated the biological relevance of such simple neuron model (Supèr et al. 2010; Supèr and Romeo 2010). Let V_{L0} , U_{L0} denote the membrane potentials and recovery variables of the L0 neurons. These quantities evolve according to laws for generic V , U given by Eqs. 1–3 (see Izhikevich 2003), namely

$$C \frac{dV}{dt} = 0.04V^2 + 5V + 140 - U + I \quad (1)$$

¹ Credit is hereby given to the Massachusetts Institute of Technology and to the Center for Biological and Computational Learning for providing the database of facial images. <http://cbcl.mit.edu/software-datasets/heisele/facerecognition-database.html>.

Fig. 1 Architecture of the spiking network. L0 is an array of spiking neurons and ON is a single spiking neuron. The arrays made of all the weights of the connections are indicated by ω , W_{L0} and W_{ON} and the term θ denotes the output threshold or bias. f is a sigmoid function

$$\frac{dU}{dt} = a(bV - U),$$

with the after-spike reset rule

$$\text{if } V \geq V_{\text{thr}} \text{ then } \begin{cases} V \leftarrow c \\ U \leftarrow U + d \end{cases} \quad (2)$$

C , V , U , I , t denote capacitance, membrane potential, recovery variable, input current and time. Further, a is a time scale, b the recovery sensitivity, c the reset value for V , and d the height of the reset jump for U . We choose $a = 0.02$, $b = 0.2$, $c = -65$, $d = 6$ (in the dimensionless convention of (Izhikevich 2003), with unit capacitance (C), and $V_{\text{thr}} = 30$ mV. These values correspond to the phasic bursting type (Izhikevich 2003). We choose the neurons to be phasic bursting because feedforward connections rely on bursting neurons, which report the beginning of the stimulation by transmitting a burst. The equations will be numerically solved by the method of (Izhikevich 2003), for an adequate value of the time step, Δt , of 0.2 ms (Euler procedure).

Input to L0 and ON in spiking model

Equations 1 particularized to layer L0 mean that the V_{L0} , U_{L0} evolutions depend on an input term like I , say I_{L0} . Let X indicate the image and $\hat{1}$ a matrix of the same size as X and made of ones. The inputs to the cells in the L0 layer are, in matrix form,

$$I_{L0} = W_{L0} \frac{X + \hat{1}}{2}. \quad (3)$$

Since the X coefficients range from -1 to 1 , the $\frac{X+\hat{1}}{2}$ coefficients will range from 0 to 1 . Our initial conditions at $t = 0$, are $V_{L0}(t = 0) = c$ and $U_{L0}(t = 0) = bV_{L0}(t = 0)$ for every cell. We set $W_{L0} = 150$, since it produced an adequate spike response.

Analogously, the output neuron (ON) is described by a membrane potential V_{ON} and a recovery variable U_{ON} . Now, the input I_{ON} to this cell will be a global factor W_{ON} times a step or sigmoid function f , applied to the weighted sum of L0 state. By applying a sigmoid we allow for some deviation from (generalization of) the initial settings. If M_{L0} is the (0,1)-spike map at layer L0, we take its ‘ $(-1, +1)$ -version’ $S_{L0} = 2M_{L0} - \hat{1}$. Recalling the ω and θ obtained after training (see section “[Training](#)”), we set,

$$I_{ON} = W_{ON} f(\omega \cdot S_{L0} - \theta). \quad (4)$$

where the indicated ‘dot-product’ has the conventional meaning. $W_{ON} = 500$ is chosen as it gives acceptable performance. We take as f a rectified hyperbolic tangent (to smooth the step function)

$$f(y) = [\tanh(\beta y)], \quad \beta = 0.05, \quad (5)$$

with [...] meaning half-wave rectification.

Training

Training was performed by means of supervised learning of images from the supplied ‘train’ sets. A linear network is trained using standard techniques (Moore–Penrose pseudo-inverse) for linear systems. In previous models we have tested STDP, which gave similar results as the standard techniques. As the STDP method requires longer running times we did not use it here. Special care is taken of the problem of over-fitting by limiting the singular values for the considered systems. First, the weight (ω) coefficients are calculated using the ‘face’ part of the training set. Afterwards, the output threshold (θ) is found using the ‘non-face’ and ‘face’ parts together. For sets of a few hundred faces and non-faces, our ω coefficients fall within the approximate range $[-0.5, 0.5]$, and θ is of the order of 10 . These values were used of testing the spiking model.

Testing the models

The system was tested on new images. Initially, the rate limit for face/non-face distinction was from 0 to 0.1 spikes/ms. In a later analysis, we varied the rate limit to determine the detectability performance of the model. This limit amounts to three spikes during 30 ms of simulation time. For three or fewer spikes the category is ‘non-face’ and, for

more, ‘face’. Spike counting was restricted to $t > 1$ ms. Images were shown for 30 ms (is period of simulation). In the backward masking experiment, images were shown for $1, 2, 3, 4, 5$ or 10 ms immediately followed by the presentation of the mask, which was created by a random permuted version of the existing image pixels. In an additional test, no mask was presented.

Human backward masking experiment

Ethics statement

Six observers with written informed consent and normal vision participated in the backward masking experiment. The study was approved by the local ethics committee (Comissió de Bioètica de la Universitat de Barcelona).

Stimuli and apparatus

We used 400 images (200 faces and 200 non-faces) from the CBCL database for the backward masking experiment. For each image a mask was created by permutation of the original image. For accurate control of timing, stimuli were presented by EventIde (a generous gift from Ilia Kournakov). Images were shown at the center of the screen on a grey background directly followed by the mask. Trial duration was 500 ms. Subjects seated in a dim light (6 cd/m²) room 30 cm from the monitor (Phillips Brilliance 22”, CRT, 100 Hz) and image size was 19 by 19 pixels corresponding to 1.4 degrees of visual angle. The display resolution was $1,024 \times 768$ pixels.

Task procedure

Subjects were tested in a two-alternative forced choice paradigm and had to indicate by a button press whether the presented stimulus was a face or non-face image. Images were briefly (10 – 100 ms) shown and followed by the presentation of the mask. Mask stimuli were presented at different times (from 10 to 100 ms in steps of 10 ms) after stimulus presentation (stimulus onset asynchronies; SOA) with an inter stimulus interval of zero. Face and non-face images and SOAs were randomly interleaved. For each SOA condition we used 80 trials (40 face and 40 non-face images).

Results

The continuous input from grey scale face/non-face images were transformed into a spike map by the first layer (L0). The layer 0 neurons produced a sustained firing pattern where spike timing differed among neurons (Fig. 2a). The output neuron that integrated all these spikes responded by a brief

burst with a variable amount of spikes, typically between 1 and 12 spikes (depending on the stimulus. Some examples of output neuron activity in the spiking model are illustrated in Fig. 2b, which shows a variety of $V_{ON}(t)$ profiles coming from our simulations. Note that the timing of the spikes: they are often elicited within the first 10 ms after stimulus presentation. For non-faces also spikes were elicited.

Model performance

We then tested the model for 400 faces versus 400 non-faces after training. We calculated the success rates (hits) for faces/non-faces (sr_f/sr_{nf}), by dividing the number of correctly classified faces/non-faces by their total numbers. Hence, the usual false alarm rate for non-faces categorized as faces, fp_f , amounts to $1 - sr_{nf}$. To obtain a reliable estimate of the performance we repeated this cycle five times. Overall, the hit rate was $70\% \pm 0.14$ and the false alarm rate was 27% ($d' = 1.1$). To test whether the face/non-face distinction is not a question of qualitatively

different power spectra we calculated the power spectra for faces and for non-faces images. The results (not shown) demonstrate that the 2D power spectra for faces and for non-faces are qualitatively similar. Therefore the classification is not a question of different frequency contents. Furthermore, the average luminance plays no role because images are rescaled to the $[-1, +1]$ range, thus taking all mean luminances close to zero.

Model responses after backward masking

The mask destroyed the space correlations to the extent their power spectra are practically flat. The power spectra for masks are similar among them and different from face or non-face power spectra. Masking resulted in the termination of the burst, i.e. there were no more spikes elicited after the presentation of the mask. Thus, the mask itself did not provoke a spike response. Consequently, the number of spikes elicited by a face/non-face image gradually decreased with shorter SOAs (Fig. 3a, b). To test whether

Fig. 2 Neural responses to face/non-face images **a** Spike timing of all layer 0 neurons. Each dot represents the time of a spike elicited after presentation of a face image. Note that responses continue during the period of stimulus presentation (≈ 30 ms). **b** Burst response of output neuron. Different examples of the evolution of the membrane potential (in mV) of the output neuron ON—called V_{ON} in the text—as a function of time, for the faces and non-faces, shown below. Time is measured from stimulus onset

the mask disrupted the target evoked bursts we measured the performance at the different stimulus durations without the mask. The results reveal (Fig. 3c) that the performance was almost identical to the mask condition. Thus, in addition to noise (Du et al. 2010) burst duration is dependent on the stimulus duration.

Model performance after masking

To analyze the performance of the model after backward masking we calculated the hit rate for the face and non-face images. These results showed that the hit rate for faces gradually declined for shorter SOAs. Performance for non-face detection (correct rejections) increased with shorter SOAs signifying that the false alarms for the face images decreased with shorter SOAs (Fig. 3d). Finally, we calculated the receiver operating characteristic (ROC) curve by changing the threshold (criterion, i.e. number of spikes) that we used to classify the responses as face or non-face. The ROC analysis showed that detectability performance remained similar across the different SOAs, except for the shortest SOAs (1 and 2 ms) when masking resulted in appreciably poorer detection performance (Fig. 4a).

Human performance of face detection after masking

Humans were tested using the same face and non-face stimuli as in the model backward masking experiment. For the shortest stimulus presentation (SOA = 10 ms) the average performance was at chance level (mean $\pm$ SEM; hits: 40% $\pm$ 0.05; false alarms: 34% $\pm$ 0.05). For longer SOAs hit rate increased ($R = 0.92$; $P < 10^{-3}$) reaching a maximum for stimulus presentation times longer than 60 ms (Fig. 4b) whereas false alarm rates tended to slightly decrease ($R = -0.64$; $P = 0.05$). The average performance (average score from the SOAs of 60–100 ms) for hits was 65% $\pm$ 0.05 and for false alarms, 31% $\pm$ 0.04 (mean $\pm$ SEM). These average responses (hit rate = 70% and false alarm rate = 27%) were similar to the model performance (see also Fig. 4a). To test for the effectiveness of masking, we repeated the experiment but without presenting the mask (in fact the mask was a black screen instead of a permuted image and is referred to as a poor mask) while leaving the stimulus duration equal. These results show that performance improved (higher hit rate and lower false alarm rate; Fig. 4b), and indicate that the mask (i.e. the permuted image) reduced hit rate and increased false alarm rate by $\sim 20\%$.

Discussion

We tested the ability of feedforward responses for face categorization. The model transformed grey scale face/

non-face images into a spike map. By integrating the spikes the output neuron responded by a brief spike burst. After training the network, the model achieved face/non-face categorization for additional sets of these images. Testing the model in a backward masking paradigm, we observed that face classification was still robust despite the strong reduction in the spike number of the output neuron.

Model versus human performance

In our model experiment we showed that despite reduced feedforward responses by backward masking, detectability performance measured by ROC analysis remained robust except for the shortest SOAs. Furthermore, our model findings show that stimulus evoked responses stopped at the moment the stimulus was removed. Thus the reduced responses after backward masking are explained by the reduction in time to integrate the sensory information. Reduced neural responses are also observed in human backward masking studies, but whether they are caused by the limited integration period remains to be tested. On the contrary, stimulus evoked responses of face selective cells survive after stimulus removal (Quian Quiroga et al. 2008). Accordingly, our model data predict that this continuation stimulus evoked responses reflect recurrent processing. This idea is in line with a recent transcranial magnetic stimulation study (de Graaf et al. 2011) that proposes an early overlap between recurrent and feedforward responses during face detection and with the observation of an early effect of feedback on feedforward responses (Hupé et al. 2001).

Feedback is known to suppress noise and to enhance stimulus contrast (Hupé et al. 1998; De Weerd et al. 1999; Huang et al. 2007; Hayes and Merigan 2007), in particular when the feedforward signal is weak (Supér and Romeo 2011). Such an idea of noise reduction by feedback may explain why in our human study the false alarm rate remained constant and even slightly decreased for longer SOAs when the role of feedback is expected to be more pronounced.

Our model neuron responded with a single burst of 1–12 spikes. The average face selective cell burst response is in the order of 8 spikes/sec in monkey IT cortex (Ó Scalaidhe et al. 1999) and 50 spikes/s in human prefrontal cortex (Quian Quiroga et al. 2008). These higher values in the human and monkey cortex can be a sign of recurrent interactions that prolong the stimulus response, e.g. into a memory component (Supér et al. 2001). For instance spike bursts of face selective cells may outlast the stimulus presentation period and can continue for several hundreds of milliseconds whereas in our model study the burst length was less than 30 ms.

According to our model experiment, the first part of the spike burst, i.e. 3 spikes is sufficient to detect the face

Fig. 3 Responses after masking. **a** Examples of spike responses after backward masking for faces and non-faces. The numbers on the left indicate the stimulus onset asynchrony (SOA). Note that the mask does not elicit spikes. Time is from stimulus presentation. Numbers on the left represent the SOAs and NM denotes no-masking condition.

b, c Average responses to faces and non-face stimuli with and without masking. Error bars represent deviation of the mean. **d** Hit and false alarm rates. Numbers on the horizontal axis represent the SOAs and NM denotes no-masking condition

Fig. 4 Receiver operating characteristic (ROC) curves and human data. **a** The ROC shows the probability of a hit versus the probability of a false alarm. Different curves are created by changing the response criterion, i.e. the number of spikes. *Numbers* in the legend represent the SOAs and NM denotes no-masking condition. Added in are the human data. *Diamonds/squares* represent the values obtained in the masking/poor masking condition. Note that only one data point per SOA condition in the human backward masking experiment can be obtained. **b** Average stimulus detection performance of the human subjects in a backward masking experiment. A hit means correct detection of a face image and false alarm an incorrect response to a non-face image, i.e. classified as a face. *Black/grey* traces represent the mask/poor mask condition. *Numbers* on the x-axis represent the SOAs

image. The interspike interval in a spike burst in the human temporal cortex is in the order of 10–25 ms (Colder et al. 1996), which means that it takes about 30–75 ms to elicit 3 spikes. These values are consistent with our results from the human masking experiment where moderate face detection was observed for stimulus presentation periods between 30 and 60 ms. Moreover, robust face detection has been found after stimulus presentation times of more than

66 ms (Quiñero Quiroga et al. 2008), possibly the time when feedback starts to kick in. Thus the difference in interspike interval may explain why in humans backward masking occurs over a much longer time scale.

Backward masking and stimulus detection

According to our model data, the reduced responses after backward masking are a result of the removal of the stimulus. Therefore, when no mask was presented after the target image performance did not change. However in the human experiment, when a poor mask, i.e. a black screen, was presented performance improved. Here the hit rate increased while the false alarm rate considerably dropped. Such a push–pull operation is ascribed to early feedback influences on the feedforward evoked responses (Hupé et al. 2001; de Graaf et al. 2011; Supér and Romeo 2011). As the stimuli and the stimulus presentation times were the same in both conditions (mask and poor mask), the improved detection performance in the poor masking condition can be attributed to feedback. Although feedback is not critical for stimulus detection, this idea supports the theory on backward masking by recurrent processing (Lamme et al. 2002; Di Lollo et al. 2000).

Acknowledgments This work was supported by grants to HS (SEJ2006-15095, PSI2010-18139 & SAF2009-10367) from the Spanish Ministry of Education and Science (MICINN) and (2009-SGR-308) from Catalan government (AGAUR).

References

- Colder BW, Wilson CL, Frysinger RC, Harper RM, Engel J Jr (1996) Interspike intervals during interictal periods in human temporal lobe epilepsy. *Brain Res* 719:96–103
- De Graaf TA, Goebel R, Sack AT (2011) Feedforward and quick recurrent processes in early visual cortex revealed by TMS? *Neuroimage*. Epub ahead of print
- De Weerd P, Peralta MR III, Desimone R, Ungerleider LG (1999) Loss of attentional stimulus selection after extrastriate cortical lesions in macaques. *Nat Neurosci* 2:753–758
- Delorme A, Thorpe SJ (2001) Face identification using one spike per neuron: resistance to image degradations. *Neural Netw* 14:795–803
- Di Lollo V, Enns JT, Rensink RA (2000) Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *J Exp Psychol Gen* 129:481–507
- Du Y, Lu Q, Wang R (2010) Using interspike intervals to quantify noise effects on spike trains in temperature encoding neurons. *Cogn Neurodyn* 4:199–206
- Hayes RD, Merigan WH (2007) Mechanisms of sensitivity loss due to visual cortex lesions in humans and macaques. *Cereb Cortex* 17:1117–1128
- Huang JY, Wang C, Dreher B (2007) The effects of reversible inactivation of postero-temporal visual cortex on neuronal activities in cat's area 17. *Brain Res* 1138:111–1128
- Hupé JM, James AC, Payne BR, Lomber SG, Girard P et al (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394:784–787

- Hupé JM, James AC, Girard P, Lomber SG, Payne BR, Bullier J (2001) Feedback connections act on the early part of the responses in monkey visual cortex. *J Neurophysiol* 85:134–145
- Izhikevich E (2003) Simple model of spiking neurons. *IEEE Trans Neural Netw* 14:1569–1572
- Lamme VAF, Zipser K, Spekreijse H (2002) Masking interrupts figure-ground signals in V1. *J Cogn Neurosci* 14:1044–1053
- Ó Scailaidhe SP, Wilson FAW, Goldman-Rakic PS (1999) Face-selective neurons during passive viewing and working memory performance of rhesus monkeys: evidence for intrinsic specialization of neuronal coding. *Cereb Cortex* 9:459–475
- Quiroga RQ, Mukamel R, Isham EA, Malach R, Fried I (2008) Human single-neuron responses at the threshold of conscious recognition. *Proc Natl Acad Sci* 105:3599–3604
- Sugase Y, Yamane S, Ueno S, Kawano K (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400:869–873
- Supèr H, Romeo A (2010) Rebound spiking as a neural mechanism for surface filling-in. *J Cogn Neurosci* 23:491–501
- Supèr H, Romeo A (2011) Feedback enhances feedforward figure-ground segmentation by changing firing mode. *PLoS One* 6(6): e21641
- Supèr H, Spekreijse H, Lamme VAF (2001) Neuronal correlate of working memory in the monkey primary visual cortex. *Science* 293:120–124
- Supèr H, Romeo A, Keil MS (2010) Feed-forward segmentation of figure-ground and assignment of border-ownership. *Plos One* 19(5):e10705
- VanRullen R, Gautrais J, Delorme A, Thorpe S (1998) Face processing using one spike per neurone. *Biosystems* 48:229–239
- Weyrauch B, Huang J, Heisele B, Blanz V (2004) Component-based Face Recognition with 3D Morphable Models, First IEEE Workshop on Face Processing in Video, Washington, DC