A MODEL OF VISUAL BACKWARD MASKING.

Guido Bugmann[1], *John G. Taylor**

School of Computing, Communications and Electronics, University of Plymouth,
Plymouth PL4 8AA, UK.
gbugmann@plymouth.ac.uk
Tel: +44 1752 23 25 66. Fax: +44 1752 23 25 40

*Department of Mathematics, Kings College London,
Strand, London WC2R 2LS, UK
john.g.taylor@kcl.ac.uk

ABSTRACT

When two successive stimuli are presented within 0 to 200ms intervals, the recognition of the first stimulus (the target) can be impaired by the second (the mask). This backward masking phenomenon has a form called metacontrast masking where the target and the mask are in close spatial proximity but not overlapping. In that case, the masking effect is strongest for interval of 60-100ms. To understand this behaviour, activity propagation in a feedforward network of leaky integrate-and fire neurons is investigated. It is found that, if neurons have a selectivity similar to that of V1 simple cells, activity decays layer after layer and ceases to propagate. To combat this, a local amplification mechanism is included in the model, using excitatory lateral connections, which turn out to support prolonged self-sustained activity. Masking is assumed to arise from local competition between representations recruited by the target and the mask. This tends to interrupt sustained firing, while prolonged retinal input tends to re-initiate it. Thus, masking causes a maximal reduction of the duration of the cortical response to the target towards the end of the retinal response. This duration exhibits the typical U-shape of the masking curve. In this model, masking does not alter the propagation of the onset of the response to the target, thus preserving response reaction times and enabling unconscious priming phenomena.

Keywords: Metacontrast masking, backward masking, visual latency, unconscious priming.

1. Introduction

Visual recognition of a target stimulus can be impaired when it is presented in close temporal proximity with another stimulus called the mask. Of particular interest is backward masking, where the mask is presented after the target. This raises several questions. For instance: 1. How can the presentation of a later stimulus affect the recognition of a stimulus that has entered the visual system earlier? 2. Why is the mask sometimes most effective when presented approximately 60-100ms after the stimulus (U-shaped behavioural response accuracy)? 3. How can the mask affect the visibility of a stimulus but not its ability to prime responses to subsequent stimuli? Masking is a peculiar phenomenon that constitutes a test for any model of primate vision.

There are two families of models of backwards masking. The first introduces a temporary store of the target information in early visual areas that supplies information for time-consuming processing in later stages. The presentation of the mask then interferes with the integrity of the

[1] To whom correspondence should be sent

store and prevents the identification of the target. The second assumes that masking information is somehow able to catch up with target information while it travels in the visual stream and is able to interfere with it.

The store model has received experimental support for instance from MEG data showing that masking corresponds to the temporal overlapping of occipital responses to the target and the mask (Rieger et al., 2002). Electrophysiological data from Rolls and Tovee (1994) show that face selective neurons in IT exhibit a prolonged firing after the offset of the target, and that this firing is interrupted by the presentation of the mask. Such prolonged firing could reflect either a local storage mechanism, or persistent input from a store in lower visual areas.

The store concept underlies the recent "efficient masking" theory (Francis, 2000) that answers question 2 by assuming that the stimulus is stored in the form of a decaying trace that is available for processing as long as its activity exceeds a given threshold. The longer the duration of the supra-threshold phase, the better the "perceptual strength" of the stimulus and the more accurate the behavioural response. The mask would reduce the trace activity and would have a maximal effect on its duration when the activity is close to the threshold, hence some time after the stimulus onset. The "object substitution" theory of Enns and DiLollo (2000) also uses a decaying image of the target in early visual areas. The mask would then not inhibit the target, but "replace" it, thus effectively interrupting its processing. All the models above are conceptual, and do not rely on details of neuronal information processing.

In contrast, in a model by Bugmann and Taylor (1994), stores took the form of sustained neuronal firing. These were integral components of each visual information processing stage, ensuring that information is kept available until recognised by neurons in the next layer. When such a scheme is modelled by a pyramidal network of spiking neurons, it is found that activity propagation is a stochastic process where activity "jumps" to the next layer at random times, and information "stays" in source neurons for a random duration. The processing is purely feed-forward, but activity propagates with variables velocities in different branches of the pyramid. Part of the stimulus information can then still be located in the first layer when the masking stimulus is presented. The mask activates lateral inhibitory mechanisms and interrupt the firing of neurons representing the stimulus. Thus, part of stimulus information can be erased by the mask, so preventing, or degrading, stimulus identification. The advantage of such a detailed model is that the probabilistic response of subjects is a feature of the model and does not need to be introduced by hand as in other trace models. It should be noted that the only role of the store is to ensure loss-less feedforward information propagation. Thus, the model is consistent with evidence that early responses in the visual stream result from purely feedforward processing (Mehta et al., 2000, Thorpe et al., 1996) and carry almost all information about the stimulus (Oram and Perret, 1992). This model answers question 1, but does not reproduce the U-shape of the response accuracy with SOA (Stimulus – Mask Onset Asynchrony).

The second family of model is exemplified by the work of Breitmeyer (1984) where an inhibitory signal is carried by the fast magnocellular pathway that allows the masking signal to "catch up" with the slower visual information in the parvocellular pathway. There are evidences for a temporary inhibition of neural response in V1 by the mask (Macknik and Livingstone, 1998), although it is unknown if the magnocellular pathway is involved. Such a model has been able to explain a number of peculiar perceptual phenomena (Breitmeyer and Ogmen, 2000) but unconscious priming (question 3) was not addressed.

Overall, the idea that the mask has the ability to interfere with ongoing activity is generally accepted. This can be in the form of more or less strong lateral inhibition due to the incompatibility of low level features of the mask and target, and sometime involves cross-stream inhibition (parvo- and magno-cellular). Target substitution could possibly also be described in these terms (Francis and Hermes, 2002). However, most current models are rather vague on how later processing stages, important for priming effects of semantic nature (Dehaene et al., 2001), are affected by the mask. In general, models where masking causes an interruption of the

response in early stages of processing will have difficulties accounting for priming effects that rely on information reaching higher visual areas.

To explore the effects of the mask on the dynamics of information propagation and recognition, the model of Bugmann and Taylor (1994) has been refined with more realistic, yet simple, models of neuronal integration, retinal persistence and memory processes along the visual information processing stream.

2. Model

2.1 Network model

Information propagation in the visual system is investigated through a 5-layer pyramidal network model where each neuron receives 4 inputs from 4 neurons in the previous layer. There is one neuron in the last layer and 256 neurons in the first (figure 1). The first layer represents the Lateral Geniculate Nucleus whose neurons essentially copy the firing of retinal ganglion cells. We call it the LGN/Retina layer. The second layer represents simple cells in the input layer 4 of the primary visual area V1. The third layer represents output cells in layer 2/3 of V1. Layer 4 and 5 represent input and output cells of the visual area V2. Four more layers would be needed to simulate the remaining layers in the temporal visual stream (V1 -> V2 -> V4 -> IT) but that would lead to excessive simulation time (65536 neurons required in the first layer) for no additional gain in understanding information propagation and masking principles.

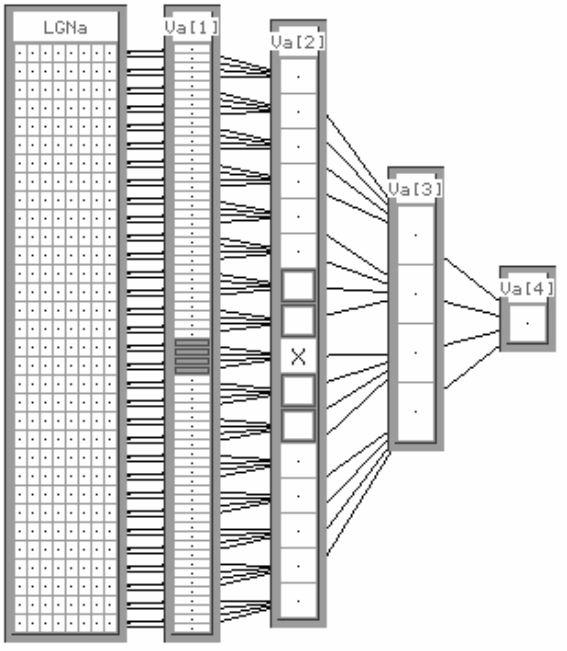

Figure 1. Pyramidal network model of the initial stages of the visual system. Each neuron has four feedforward and lateral excitatory inputs. The accentuated cells in the third layer (Va[2]) are the 4 lateral inputs to the cell marked with an X, and the 4 small accentuated cells in the second layer (Va[1]) are its four feedforward inputs. The neuron in the last layer has 5 self-feedback connections instead of lateral inputs.

This pyramidal architecture is a very simplified model of cortical connectivity. It represents the subnet of neurons that contribute to the response of one neuron in a higher layer. It complies with

the standard view that the visual system is hierarchical and convergent (see e.g. Rolls & Milward, 2000). The number of inputs to each neuron is much smaller than in reality but this is compensated by the stronger weights of each input. Setting the same number of inputs for each neuron allows using the same neuronal parameters throughout the network.

2.2 Neuron model

i) LGN/retina neurons. The output of the retina is made of spikes produced by retinal ganglion cells. These spikes are relayed by the Lateral Geniculate Nucleus (LGN) to the primary visual area V1 without significant alteration of their temporal properties. Thus, retina and LGN are aggregated here into one type of neuron. In response to brief flashes of light, comparable to the brief stimulus durations considered here, ganglion cells produce prolonged spike trains with a minimal duration of around 40-80ms, depending on the brightness and contrast of the stimulus (Levick and Zack, 1970). Under different conditions, more complex response patterns, with pauses and rebounds, are observed (see e.g. Schiller, 1969), but there is no need to model these here. The firing probability of ganglion cells is thus modelled as a single alpha function with a prolonged middle plateau having the duration of the stimulus:

$$f = \frac{ef_{\max}}{t_{\max}}(t - \Delta t_{onset})\exp(-\frac{t - \Delta t_{onset}}{t_{\max}}) \qquad if \quad t < t_{\max} + \Delta t_{onset}$$

$$f = f_{\max} \qquad if \quad t_{\max} + \Delta t_{onset} < t < t_{\max} + \Delta t_{onset} + \Delta t_{stim}$$

$$f = \frac{ef_{\max}}{t_{\max}}(t - \Delta t_{onset} - \Delta t_{stim})\exp(-\frac{t - \Delta t_{onset} - \Delta t_{stim}}{t_{\max}}) \qquad if \quad t > t_{\max} + \Delta t_{onset} + \Delta t_{stim}$$

Where e is exp(1). f_{max} is the maximum firing rate at the peak of the alpha function. In most simulations f_{max} =100Hz. Δt_{onset} is the onset delay of the retinal response. We used Δt_{onset} =25ms from figure 10a in Bair et al., (2002). Δt_{stim} is the duration of the stimulus, taken here to be 10ms. t_{max} is the time at which the alpha function reaches its maximal value, corresponding to approximately ¼ of the minimal duration of the retinal response set here to 60ms (i.e. t_{max} = 15ms).

ii) Neurons in the pyramidal network. As a model for all neurons in the network we used a LIF neuron model tuned to reproduce the selectivity of simple cells in V1. It can be estimated from orientation selectivity data that simple cells in V1 have a relatively high selectivity, firing only when more than 75% of inputs are active. Thus, using 4 inputs to a neuron is the absolute minimum necessary to express this selectivity, i.e. the LIF model used here is tuned to fire for 4 active inputs and be barely producing spikes with 3 inputs.

Tuning stages are as follows. First we selected the best parameters for high selectivity found in Bugmann (2002): These are a somatic decay time constant of 100ms, depressing synapses with a recovery time constant of 100ms, partial reset of the membrane potential after a spike to 90% of the firing threshold of 15mV, and EPSCs modelled as alpha functions with a maximum amplitude at 1ms after spike arrival. Then, we set 3 active inputs to fire random spikes at 100Hz and we searched by trial and error the synaptic weight that sets the neuron just at the limit between firing and not firing. The value found was 1.87, which is the maximum of the current described by the EPSC alpha function (see Bugmann et al., 1997 and Bugmann, 2002 for translating such values into conductances). The resulting model, although not incorporating many physiological details, produces output spike trains with firing rates, coefficients of variations and contrast-invariant tuning curves comparable to those of neurons in V1. It may not be correct to use the same model in layers corresponding to higher visual areas, but there are no data on their input-output functions which could inform the design of more specific models.

2.3 Need for a local amplification mechanisms.

The first experiment performed was to generate sustained 100 Hz firing in the LGN/retina and observe if activity propagated through the network to the last layer. Neurons only had feedforward connections. It can be seen in figure 2A that the activity decreases in successive layers and barely reaches the fifth layer. The reason is that, due to the high selectivity of the neuron, its output firing rate is lower than the input firing rate. Despite synaptic depression which tends to equalize post-synaptic currents generated by a range of input firing rates, some of the input rate reduction is reflected in the output rate (see figure 8 in Bugmann (2002)). Considering that the visual system has several more layers before IT, the model would produce no response in IT, contrary to observations showing strong responses (see e.g. Oram and Perret, 1992).

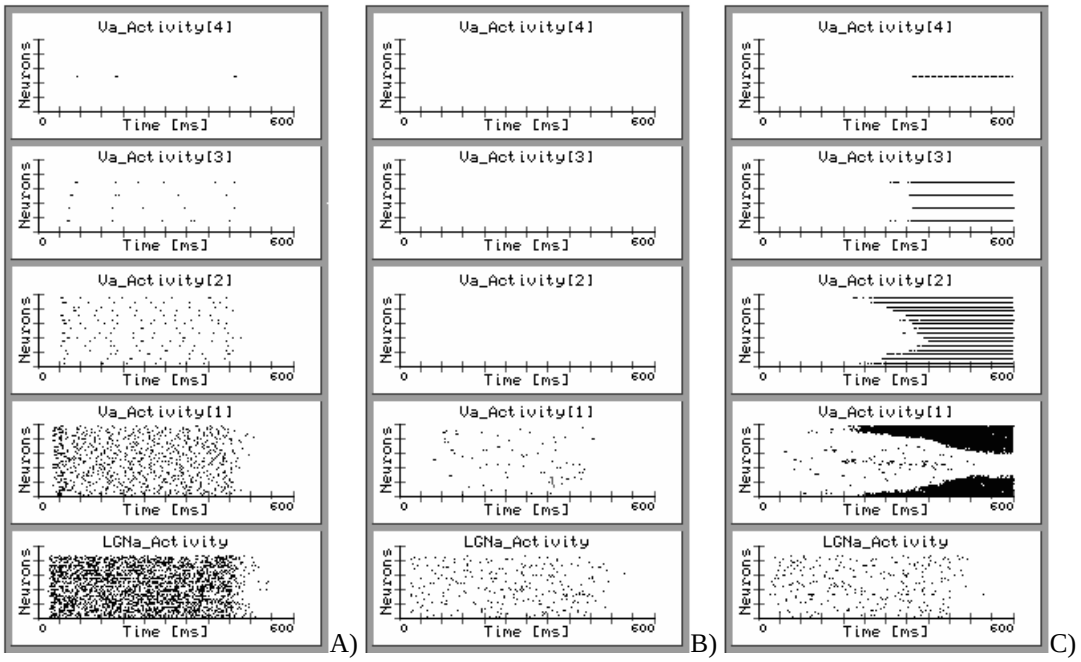

Figure 2. Activity propagation in the network. The graphs show spikes produced by the neurons in successive layers in a single run. A and B. Without local amplification mechanism and retinal input at 100Hz and at 10Hz respectively. C. With a minimal gain in each layer generated by excitatory lateral connections. The retinal input fires at 10Hz, representing a stimulus lasting 400ms.

Figures 2A and 2B indicate that some form of local amplification mechanisms is needed to ensure information propagation in the visual system. There is evidence in Chung and Ferster (1998) that cortical feedback approximately doubles or triples the input current to a neuron in a proportional way. A source of background input is suggested in (McCormick et al., 2003) to explain sustained firing, possibly linked to attentional modulation. Other observations such as waves travelling along the cortical surface (Prechtl et al., 1997) also indicate the presence of mechanisms strong enough to sustain activity. The tuning width of the cortically generated currents reported in Chung and Ferster (1998), and the fact that they are eliminated by a suppression of local cortical firing, suggests that these are generated by spikes from neighbouring neurons with slightly different tuning properties that are also activated by the stimulus. In our simplified network, such amplifying input is simulated in the form of lateral connections received from the two

neighbouring nodes on each side[2] (figure 1), with a circular topology at the edges. The last layer is provided with five self-feedback connections with the same properties as lateral connections.

The lateral weights were adjusted by setting a continuous (random) input firing rate of 10 Hz in the retina/LGN (which robustly generates response in the next layer, but not beyond (figure 2B)). Lateral weights were non-depressing (Gil et al., 1997) and had 53% probability of transferring an incoming spike and had propagation latencies between 1 and 5 milliseconds. Then the lateral weight values were increased until the last layer shows firing. Interestingly, the required weights were so large (0.45) that a runaway activity tended to develop (figure 2C). During subsequent simulations, the lateral weights were slightly reduced to allow amplification but without runaway behaviour (0.38 in the second layer, 0.33 in the third layer and 0.31 in the fourth layer). These weights produced robust propagation with the inputs at 100Hz used for masking studies. The weights needed to be smaller in higher layers because runaway was more likely due to their smaller number of neurons. Note that the model is quite sensitive to the gain in the lateral connections because the inhibitory activity control mechanism in the cortex has not been implemented. This also explains the high firing rates that the model produces in cortical layers (figure 3).

2.4 Simulation of the effect of the mask.

In our previous model, masking was modelled as a suppression of retinal/LGN activity by the mask. This generated monotonical psychometric curves (Bugmann and Taylor, 1994) but was not able to reproduce the U-shaped response curves observed with metacontrast masking (Schiller, 1969). Metacontrast masking is likely to operate at the cortical level (Schiller, 1969). Electrophysiological data show that the mask temporarily depresses the activity of cortical cells (Macknik and Livingstone, 1998). Thus, we simulate the effect of the mask as an instantaneous reduction of the membrane potentials of all cells in the second layer (V1) to 65% of their previous values. This occurs 25ms after mask presentation, where 25ms is the retinal latency (see 2.2.i). This simulation of masking effects is extremely simplified and a more detailed modelling of cortical dynamics would be required in a quantitative model, for instance along the lines developed in Wang et al., (2004). The assumption here is that the mask interferes with the stimulus through a competition between incompatible representations. An example of such process is cross-orientation inhibition. Masking is an interference between propagating activity representing the mask and activity representing the target, in two different subnets of the visual system located in close proximity. In principle, interference could also affect higher layers in our model, but for a demonstration of the mechanism, only interference in V1 is considered.

3. Results

We presented a 10 ms target stimulus, resulting in a retinal response lasting 70-80 ms. We then presented the mask with a delay between 0 and 200ms. The puzzling aspect of the observed effect on activity propagation is that, at no time did the mask prevent activity from reaching the last layer (Fig 3B and 3C). The mask however reduced the duration of firing in all

[2] One may argue that, as this network only represents a subnet of the visual system relevant for a given stimulus in a given position, neighbors in the model may represent neurons physically distant from each others. Therefore, it may be formally incorrect to model lateral inputs in the way. Fortunately, this has no consequences for the conclusions of this paper.

layers. A similar observation was made in the temporal visual cortex of the macaque (Rolls and Tovee, 1994).

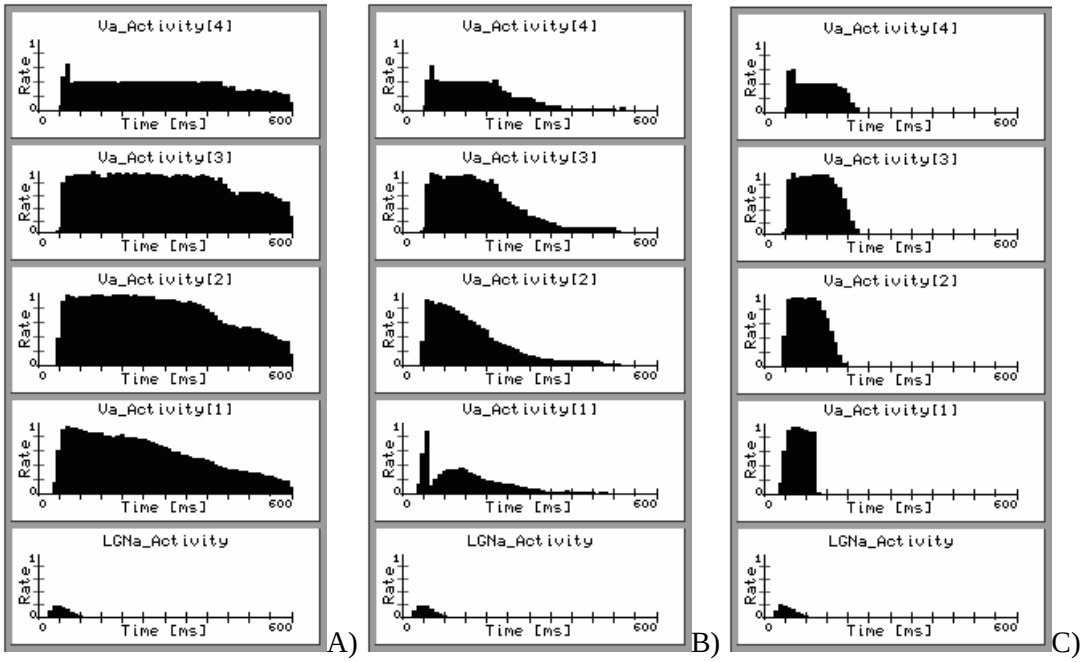

Figure 3. Histograms of activities in successive layers during masking. A. No mask applied. B. Mask applied 40ms after stimulus onset. C. Mask applied 100ms after stimulus onset. A rate of 1 on the vertical scale corresponds to 500 Hz.

When we plotted the duration of the firing in the last layer in dependence of the time difference between the target and the mask onset (SOA), we found a U-shaped relation (figure 4).

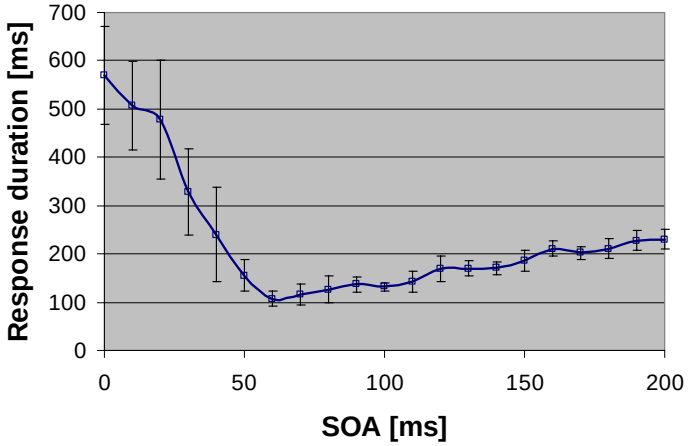

Figure 4. Relation between the response duration in the last layer (Va[4]) and the stimulus–mask onset asynchrony (SOA). Each data point is the average duration at 1/3 of the peak amplitude of the post-stimulus histogram in the last layer (figure 3). The error bars represent one standard deviation over 10 trials.

The U-shape can be explained as follows. As long as cortical cells receive retinal/LGN inputs, the resetting of the sustained cortical firing by the mask has only a temporary effect and cortical cells are able to resume sending information to the next layer (e.g. Fig. 3B). However, towards the end of the retinal response, the input is weaker and loses its ability to re-initiate cortical firing (e.g. Fig. 3C). That is when the masking effect is maximal. For later presentations of the mask, the mask simply interrupts the ongoing sustained firing and a linear relation between SOA and response duration can be expected. Indeed, such a U-shape cannot be obtained if the mask interrupts directly the retinal firing. In metacontrast masking this is avoided by having non-overlapping stimulus and mask. Thus, information about both stimulus and mask can propagate into higher visual areas in different subnets. It is only when they recruit competing representations in a given layer that interruptive masking effects are expected to occur. Thus, depending on the mask characteristics, the model produces a monotonic or U-shaped masking function.

In this model, unconscious priming is possible, as the mask does not prevent information about the target to propagate into the visual system. It reduces the duration of the response, a fact that could explain weakening imaging signals observed by Dehaene et al. (2001) and Vogel et al. (1998).

An implication of the results is that the duration of the response must relate to the visibility of the stimulus. For instance, in Schiller (1969) it was observed that the U-shape represents the perceived brightness of the masked stimulus. If we assume that the activity in the last layer is integrated by another brain area to produce the behavioural response, then that area could determine the brightness of the stimulus by comparing the duration of the firing with some standard. A similar relation was assumed in the model of Ogmen et al. (2003) where the perceived stimulus brightness was defined as the integrated response of cortical cells. One may note that, with depressing synapses as used here, postsynaptic currents do not allow a good discrimination between input firing rates. Thus, the integrated effect of an input is almost exclusively related to its duration.

The fact that reaction times are not affected by the mask (Schiller, 1969) can be explained here by the fact that the propagation velocity of the onset of the response is not affected by the mask.

Other authors have suggested that conscious perception is related to the integrity of the tail of the neuronal response, i.e. 100 ms and later after response onset (Mehta et al, 2000; Schiller, 1969). This may be linked to the presence of an attentional feedback signal propagating from higher layers to lower layers (Mehta et al., 2000). In that case, the width and timing of the bottom of the U-shape should be related to where in the visual hierarchy the feedforward masking signal meets with the feedback attentional signal. Future experiments should be able to test this hypothesis. Thus, response duration alone may not be the main factor in stimulus discrimination and there may be some special function supported by the later part of the response which might be disrupted by masking.

4. Conclusion

In summary, the effect of the mask on activity propagation in a very simplified model of the visual system is investigated. In this model, the response to the mask interferes with sustained cortical responses to the target maintained by a local excitatory recurrent network. Sustained cortical responses are supported by prolonged retinal responses, and the mask only becomes effective at the end of the retinal response. Its main effect is to reduce the duration of the sustained response to the target. Such a model produces U-shaped behavioural responses if one takes the duration of the response is an indicator of stimulus visibility.

A testable prediction of the model is the relation between the minimum of the U-shaped curve and the duration of the retinal response. As the latter depends on stimulus intensity or contrast, the characteristics of the U-shape might depend on stimulus intensity or contrast.

As the mask only affects the tail of the response, it is suggested that unconscious priming may be mediated by the early components of the response.

Acknowledgement

The authors a grateful for comments by anonymous referees who helped improve this paper.

References

[1] Bair W., Cavanaugh J.R., Smith M.A. and Movshon J.A. 2002. The timing of response onset an offset in macaque visual neurons. *J. Neuroscience*. 22(8):3189-3205.

[2] Breitmeyer, B. G. 1984. Visual masking: An integrative approach. *New York: Oxford University Press.*

[3] Breitmeyer B.G., Ogmen H. 2000. Recent Models and Findings in Backward Visual Masking: A Comparison, Review, and Update', *Perception & Psychophysics*, vol. 62, pp. 1572-1595.

[4] Bugmann G. and Taylor J.G. 1994. Role of short-term memory in visual information propagation. Extended Abstract Book of Int. Symp. on Dynamics of Neural Processing, Washington, 132-136

[5] Bugmann G. 2002. Synaptic depression increases the selectivity of a neuron to its preferred pattern and binarizes the neural code. *Biosystems: Special Issue on Neural Coding*, 67, 17-26.

[6] Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J. F., Poline, J. B., & Riviere, D. 2001. Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, 4, 752-758.

[7] Enns, J.T., DiLollo, V. 2000. What's new in visual masking. *Trends in Cognitive Sciences, 4*, 345-352.

[8] Francis, G. 2000. Quantitative theories of metacontrast masking. *Psychological Review, 107*, 768-785.

[9] Francis G. and Hermens F. 2002. Comment on “Competition for consciousness among visual events: the psychophysics of reentrant visual processes” (Di Lollo, Enns, & Rensink, 2000) *J. Exp. Psychology: General*, 131(4): 590-593

[10] Gil Z., Connors B.W. and Amitai Y. 1997. Differential regulation of neocortical synapses by neuromodulators and activity. *Neuron.*19:679-686.

[11] Levick W.R. & Zack J.L. 1970. Response of cat retinal cells to brief flashes of light. *J. Physiol*. 206, pp. 677-700.

[12] Macknik S.L. and Livingstone M.S. 1998. Neuronal correlates of visibility and invisibility in the primate visual system. *Nature Neuroscience* 1(2), 144-149.

[13] McCormick D.A., Shu Y., Hausenstaub A., Sanchez-Vives M., Badoual M. and Bal T. 2003. Persistent cortical activity: Mechanisms of Generation and Effects on neuronal modulation. *Cortex*. 13:1219-1231.

[14] Mehta S.D., Ulbert I., Schroeder C.E. 2000. Intermodal selective attention in monkeys. I: Distribution and timing of effects across visual areas, *Cerebral Cortex*, 10 (4): 343-358.

[15] Oram, M.W. & Perrett D.I. 1992. Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology , 68:70-84.*

[16] Prechtl J.C., Cohen L.B., Pesaran B., Mitra P.P. and Kleinfeld D. 1997. Visual stimuli induce waves of activity in turtle cortex. Proc. Natl. Acad. Sci. USA, 94(14):7621-7626.

[17] Rieger, J.W., Braun C., Gegenfurtner K.R. and Bülthoff H.H. 2002. The Dynamics of Visual Pattern Masking in Natural Scene processing: A MEG-Study. *(Technical report no. 103)*, Max Planck Institute for Biological Cybernetics, Tübingen, Germany (oct 2002)

[18] Rolls,E.T. and Milward,T. 2000. A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural Computation 12 (11):* 2547-2572

[19] Rolls E.T. and Tovee M.J. 1994. Processing speed in the cerebral cortex and the neurophysiology of visual masking. Proc. R. Soc. Lond. B. 257:9-15.

[20] Schiller P.H. 1969. Behavioural and electrophysiological studies of visual masking. In Leibovic K.N. (ed) Information processing in the nervous system. Springer. P.141-165.

[21] Thorpe S., Fize D., and Marlot C. 1996. Speed of processing in the human visual system. *Nature*, 81, 520-522.

[22] Vogel EK, Luck SJ & Shapiro K.L. 1998. Electrophysiological Evidence for a Postperceptual Locus of Suppression During the Attentional Blink. J Exp Psychology 24, 1656-1674

[23] Wang X.-J., Tegnér J. , Constantinidis C. and Goldman-Rakic P.S. 2004. Division of labor among distinct subtypes of inhibitory neurons in a cortical microcircuit of working memory, Proc. Natl. Acad. Science USA, 101, 1368-1373

Word count:

- abstract 275->221
- body of text 3182 -> 3592
- word equivalent of figures: 810 -> 880 (fig 1 –> 1col; fig 2,3 –> 2 col; fig 4 –> 1 col)
- reference equiv = 23x21=483
- Total: 4750 -> 5176