## Cortical correlate of pattern backward masking

(inferior temporal cortex/temporal integration/shape recognition)

GYULA KOVÁCS, RUFIN VOGELS\*, AND GUY A. ORBAN

Laboratorium voor Neuro- en Psychofysiologie, Katholieke Universiteit te Leuven, Campus Gasthuisberg, B-3000 Leuven, Belgium

Communicated by James M. Sprague, University of Pennsylvania School of Medicine, Philadelphia, PA, March 20, 1995 (received for review January 23, 1995)

ABSTRACT The perception of a briefly presented shape is strongly impaired when it is followed by another pattern, a phenomenon called backward masking. We found that the vast majority of a sample of shape-selective neurons in the macaque inferior temporal cortex respond selectively to backward-masked shapes, although these shapes could not be discriminated by human and monkey subjects. However, this selective response was brief, since it was either interrupted by the mask or overridden by a response to the mask itself. We show that reliable discrimination of briefly presented shapes by single neurons depends on the temporal integration of the response. Presentation of the mask, however, reduces the number of spikes available for integration, explaining backward masking. These results also provide direct neurophysiological evidence for the "interruption theory" of backward masking.

The detection and identification of a stimulus (target) can be impaired by the presentation of another stimulus (the mask). It is well known that this masking also occurs when the mask is presented after the target (1). Several types of this backward masking exist (reviewed in ref. 2), depending on the kind of stimuli used (light flashes versus patterns), the spatial overlap of target and mask, and the function relating the degree of masking to the temporal interval separating target and mask onset [stimulus onset asynchrony (SOA)]. These counterintuitive effects have given rise to considerable philosophical interest (3), and several theories (see ref. 2) have been proposed concerning their origin, ranging from neurophysiological models (4) to more cognitive ones (5).

In the present experiments, we studied one type of backward masking in which the targets and the mask were patterned stimuli (shapes), presented at the same retinal location, and in which the target identification decreases monotonically with decreasing SOA. This pattern backward-masking paradigm has become a widely used psychophysical paradigm in the investigation of human visual perceptual processes (e.g., see refs. 6-8), but there is no direct information available regarding the effect of the mask on the responses of the cortical neurons underlying these visual processes. Indeed, single cell recording studies on backward masking so far have used masking by light instead of by pattern (reviewed in ref. 9) or presented target and mask at different locations [metacontrast masking (10-12)]. Both of these types of masking differ from the one studied here. Masking by mere luminance stimuli is largely monocular, presumably retinal in origin (9). In contrast, the dependence of pattern backward masking on the similarity of target shape and mask (13, 14), as well as its interocular transfer (1, 15), suggests a cortical origin. Metacontrast implies that target and mask do not overlap spatially and frequently shows nonmonotonic functions relating SOA and target interference (2), unlike the pattern backward masking we investigated.

To understand the cortical mechanisms underlying pattern backward masking, we recorded the activity of single inferior temporal cortex (IT) neurons stimulated with briefly presented shapes, followed or not followed by a mask. It is known that single units in IT show strong shape selectivity, a property that may underlie shape recognition (16–18). Since we were interested in studying the neuronal mechanisms of backward masking by relatively complex form stimuli, we recorded from shape-selective cells in IT. Furthermore, IT responses can be affected by backward masking, whether the effect arises in IT itself or at earlier levels, since it is the last unimodal visual area of the ventral stream in the monkey (19, 20). Thus, IT is an interesting starting point for a neurophysiological study of pattern backward masking.

It has been demonstrated that rhesus monkeys behaviorally show backward masking by using a uniform light flash as mask (21). Since, as we noted above, masking by light differs from masking by patterned stimuli, and we wish to correlate perceptual performance and neuronal response properties, we also tested psychophysically whether rhesus monkeys show backward masking by using the same stimulus conditions as in the physiological experiments. The same animals served as subjects in both psychophysical and physiological experiments. Thus, we measured the shape selectivity of single neurons under conditions that produce backward masking psychophysically in the monkeys and determined the effect of the mask not only on the response level but also on the shape discrimination capability of the neuron. Indeed, it is the capacity of the neurons to respond differentially to shapes and not the response level per se that is the critical parameter determining psychophysical discrimination performance.

## **MATERIALS AND METHODS**

**Subjects.** One naive subject and G.K. participated in the human psychophysical experiments. Two juvenile male rhesus monkeys (*Macaca mulatta*) were trained and tested in the psychophysical experiments and served as subjects for the electrophysiological experiments. During the experiment, the monkeys were water-deprived but had dry food ad libitum. Supplementary water was given to the monkeys when necessary. The procedures conformed to guidelines established by the National Institutes of Health for the care and use of laboratory animals. After these and other experiments, the monkeys were killed with an overdose of barbiturate and then perfused. Brain sections were stained for myelin and cresyl violet. Histological analysis showed that in both monkeys we recorded in IT. More specifically, penetrations were found to be in the midpart of area TE (22) in one monkey and in anterior TEO/posterior TE in the other monkey. Thus, the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: IT, inferior temporal cortex; PSTH, peristimulus time histogram; ROC, receiver operating characteristic; SOA, stimulus onset asynchrony.

<sup>\*</sup>To whom reprint requests should be addressed.

recordings spanned a wide range of anterior/posterior positions in IT cortex.

**Apparatus.** Stimuli were generated on a Barco (Kortrijk, Belgium) CD233 monitor (frame rate, 50 Hz; P22 phosphor) using dedicated hardware developed at the Biophysics Institute (J. J. Koenderink) of Utrecht, Utrecht, The Netherlands (see ref. 23). The background luminance in the experimental room was 1 cd/m<sup>2</sup>. Single cell recordings were made with parylene-coated tungsten microelectrodes as described (23, 24). In the monkey experiments, eye movements were monitored with the scleral search coil technique (25).

Stimuli. The stimuli consisted of shape outlines (width of the outlines, 15 arcmin; contrast, 95%; approximate enclosed area, 9 degree<sup>2</sup>), presented on a textured background consisting of a random dot pattern with a 50% density and a dot size of 5 arcmin, which was replaced on every trial. The average luminance and contrast of the background pattern were 8 cd/m<sup>2</sup> and 95%, respectively. The discriminative stimulus consisted of a grating and a star-like shape for one human subject and for one monkey, while an "H" and a grating served as target for the other subjects. These shapes were chosen from a standard set of eight shapes used previously (23). All the shapes were composed of simple features such as oriented line segments, so that several shapes shared elementary features. The mask was a pattern consisting of the components of the eight shapes (see Fig. 1 *Inset*) and spatially overlapped the target.

Experimental Procedures. Psychophysics. Both human subjects and monkeys had to discriminate two shapes within each session. Subjects indicated their response either by pushing one of two keys (humans: average number of trials per condition, 400) or by a leftward or rightward saccadic eye movement (monkeys: average number of trials per condition, 1600). Both humans and monkeys were required to fixate a spot during the time period beginning 700 ms before stimulus presentation and ending at stimulus offset. In a single trial, one of the targets was presented foveally against the background pattern. In masked conditions, the mask was presented for a period lasting 300 ms minus the target duration and immediately (i.e., on the next frame) followed the offset of the target shape. Although mask duration covaried with target duration, only small differences (maximal, 14%) in mask duration were present in the critical conditions, since the masking effect occurred only for SOAs shorter than 60 ms, and therefore could not have affected the results. Correct responses by the monkeys were rewarded by drops of apple juice, but no feedback was given in the human psychophysical experiments.

*Electrophysiology.* An initial test determined the selectivity of the neurons for the standard eight shapes presented for 300 ms. If the unit was judged to be shape selective after examination of the peristimulus time histograms (PSTHs) on line, responses to the "best" shape (i.e., the one eliciting the strongest response by the neuron) and "worst" shape (i.e., the one eliciting no or the weakest response) were determined at 20-, 40-, 80-, and 160-ms target durations with and without the mask (the same mask used in the psychophysical experiments). These 16 conditions were presented in an interleaved fashion with a minimum of 10 trials per condition (mean, 15 trials), excluding those trials in which the monkey broke fixation during stimulus presentation.

**Data Analysis.** Off-line, we computed the average net number of spikes within the first 20, 40, 60, and 160 ms of the response for each unit, at each of the four target durations, taking into account the response latency of the cell in the unmasked conditions. The response latencies of the cells were determined with a variant of the CUSUM technique (26) using the 160-ms-duration conditions. The average net response of a particular condition was calculated by subtracting the number of spikes in a given bin during fixation, prior to target onset, from the number of spikes in a bin of the same duration after response onset. We used ANOVA to test the significance of the responses to a given shape and the significance of shape selectivity. Tests were classified as significant if the corresponding type I error was < 0.05.

## **RESULTS AND DISCUSSION**

To verify that rhesus monkeys show pattern backward masking, we trained the two animals used in the recording sessions to discriminate shapes at short durations. The results shown in Fig. 1 indicate that rhesus monkeys, like humans, show backward masking under our experimental conditions; displaying the mask after brief target presentations considerably impaired the shape discrimination in both species. The effect of the mask is quite strong at 20-ms target duration and decreases with increasing target duration, a finding quantitatively similar to previous human studies of pattern masking (15).

In the electrophysiological experiment, we measured responses of shape-selective cells to two targets, selected in preliminary tests of each neuron as the best and worst stimulus at four target durations with or without mask. In the analysis of this experiment, we accepted only those neurons that (i)were shown to have a statistically significant effect of shape (ANOVA; P < 0.05) in the initial test with eight shapes at 300-ms target duration, and (ii) that responded in the 20-ms unmasked condition, as judged from inspection of the PSTH. Fifty-four neurons met both criteria and Fig. 2 represents the population PSTH of these neurons, for the best and worst shapes at the four target durations in masked and unmasked conditions. For this figure, PSTHs were normalized with respect to the highest bin count (bin width, 20 ms) for each neuron and then these normalized PSTHs were averaged over cells. PSTHs were aligned according to target presentation onset, not to neuronal response onset. Thus, the width of the activity also reflects differences in response latency between neurons. Since we are interested in the effect of the mask on the shape selectivity rather than on the response per se, we subtracted the responses to the worst shapes from those of the best shapes (rows B-W in Fig. 2). The differences in average response levels for the best and worst shapes were statistically significant (Scheffe test; P < 0.05) at all target durations in both masked and unmasked conditions. In particular, this analysis indicates that, on average, shape-selective responses occur even at the 20-ms masked target duration (Scheffe test; P < 0.0003), a condition in which the masked shapes could be



FIG. 1. Results of the psychophysical experiments demonstrating backward masking in humans and in rhesus monkeys using the same stimuli as in the single unit recordings. Average proportion of correct responses is plotted as a function of the target duration for two trained human subjects (open symbols) and for two monkeys (solid symbols). Ends of range bars indicate scores of the subjects for 20-ms target duration. Solid lines, data when the target was followed immediately by another shape (mask in *Inset*, masked condition); dashed lines, data obtained when only the target was present (unmasked condition).



FIG. 2. Average of the normalized PSTHs of 54 shape-selective IT neurons at 20-, 40-, 80-, and 160-ms target durations. Target and mask presentations are indicated by solid and open horizontal bars below each histogram, respectively. Rows 3 and 6 show difference between the average PSTHs (B–W) of the best and worst targets for the masked (M) and unmasked (UM) conditions, respectively.

discriminated with an accuracy of only 65%, which is close to chance (50%) and well below conventional defined threshold levels (75%) for either human observers or monkeys (see Fig. 1). In fact, the average number of spikes in the first 20-ms period of the response (i.e., number of spikes in a bin starting at response onset with a width of 20 ms) was not significantly different in the masked and unmasked conditions at 20-ms target duration [main effect of mask,  $F_{(1,53)} = 1.13$ , not significant; interaction mask × shape,  $F_{(1,53)} = 0.1$ , not significant; mean number of spikes (±SE) for the best and worst shapes in the masked condition,  $0.65 \pm 0.09$  and  $0.23 \pm 0.06$ , respectively; unmasked condition,  $0.59 \pm 0.07$  and  $0.17 \pm 0.04$  spike].

On average, the mask did not reduce the initial response to the target but shortened it. This could be shown by analyzing the responses in the 20-ms-duration conditions with different bin widths or temporal integration times. A longer response will give increasingly larger numbers of spikes in bins with increasing widths than a shorter response. Indeed, the differences between the best and worst responses for the 20-ms unmasked condition increased significantly when longer bin widths were used to quantify the response-i.e., with longer integration times (Fig. 3A, dashed line). This indicates that the shape-selective response persisted much longer than the 20-ms presentation time. However, in the case of the 20-ms masked condition, the shape-selective response did not increase significantly with increasing integration time (Fig. 3A, solid line), showing that the masked response did not last much longer than the 20-ms presentation time. The reason for this brief shape-selective response in the case of masking was that either the unit started to respond to the mask, thus increasing its responses in the worst target condition as well (Fig. 4A), or that the response to the best shape was suppressed by the mask, even in cells not activated by the mask itself (Fig. 4B). The latter is an important result since it provides direct physiological evidence for the "interruption theory" of visual masking (see ref. 27), which states that processing of the stimulus is interrupted by the mask, thus limiting the information available for further processing. This theory is usually contrasted with the "integration theory" (9, 27), which explains visual masking as the result of temporal integration of the responses to mask and target into a single response that is unspecific for the target. Evidence for this integration, which is present as early as the retinal level, has been obtained in physiological studies of masking by light flashes (9). Our results (Fig. 4B), however, indicate that in the cortex the mask halts the response of the neuron, even in those neurons that do not respond to the mask itself.

A very recent study (28, 29), published during the preparation of this report (30), showed that neurons in the superior temporal sulcus also responded briefly to a face presented for 16 ms and followed by a mask. Our results show that the backward masking has the same effect for shapes in more ventral parts of the temporal cortex. More importantly, our data indicate that the neurons not only responded to these briefly presented, masked stimuli but remained selective for their shape during the brief early response period.

The presentation of the mask after 20 ms of target duration limited the information available for shape discrimination to



FIG. 3. Difference in response level between best and worst shape as a function of integration time. (A) Target duration of 20 ms. (B) Target duration of 40 ms. Solid line, masked condition; dashed line, unmasked condition. Symbols represent averages of the same 54 neurons as in Fig. 2. Vertical bars represent SEM.



FIG. 4. Average net neuronal response in the 20-ms masked (*Left*) and unmasked (*Right*) conditions as a function of integration time. (A) Mean number of spikes for those units (n = 25) that were activated by the mask as determined by inspection of the PSTHs of the masked worst conditions. (B) Mean number of spikes for neurons (n = 13) not activated by the mask. Those cells (n = 16) showing an unclear or only a small response to the mask were excluded. Note that even when the neurons are not activated by the mask, the shape selectivity does not increase with longer bin widths in the masked condition (B, *Left*) as it does in the unmasked condition (B, *Right*), indicating that even for these units the response is interrupted by the mask. Solid symbols, best shape; open symbols, worst shape.

the initial 20-40 ms of the response. How reliable are these brief selective responses at the single unit level? In the 20-ms masked and unmasked conditions, the average difference between the best and worst shapes, taking into account only the first 20 ms of the responses, was only 0.4 spike. This small difference is unlikely to be sufficient for reliable shape discrimination. We examined this reliability issue further by performing a receiver operating characteristic (ROC) analysis (26, 31) on the responses in the 20-ms unmasked condition for 10 units, samples for at least 40 trials per condition. The ROC analysis showed that an ideal observer, using only the spike counts of the initial 20-ms bin, can discriminate the unmasked shapes presented for 20 ms with less accuracy (median, 74%) correct; range, 53-86%) than can be done behaviorally. This indicates that the initial 20 ms of the response of a single IT cell is not sufficiently reliable to produce accurate shape discrimination and suggests that later phases of the response have to be taken into account in order to obtain reliable shape discrimination at the single neuron level. Indeed, extending the temporal window in which spike counts are integrated increases the discrimination capacity of the neurons in the 20-ms unmasked condition [ROC analysis of the same data with a bin width of 80 ms: median, 83% correct; range, 65-99%; difference with results of 20-ms bin statistically significant (Wilcoxon matched pair test; T = 0; P < 0.005)], in agreement with previous studies (26, 32, 33). Thus, lengthening the time period over which one integrates the response of the neuron results in a discrimination accuracy in some single units as good as that observed psychophysically in the unmasked condition. In the

masked condition, the use of longer integration times will not improve the discrimination capacity since, as shown in our main experiment, the neuron's shape-selective response is interrupted by the mask and does not last much longer than 20 ms.

These results suggest that the brief early shape-selective response is not sufficiently reliable to support accurate shape discrimination, explaining the effect of backward masking. One could argue that pooling the output of many of these single units can increase the discrimination capacity of that ensemble of neurons considerably (34). However, this increase in signal/noise ratio in the population response is limited by the degree of correlation of the responses of the individual units (32, 35). The response correlations of adjacent neurons have been measured by Gawne and Richmond (36) in IT and are, on average, similar to those reported in a more extensive study of correlated neuronal discharges in macaque visual area MT (medial temporal) (35). Both studies obtained low response correlations (Pearson coefficient,  $r \approx 0.20$  for units with similar stimulus preferences), but these are nonetheless sufficient to curtail the beneficial effects of pooling (32, 35). Thus, assuming that the types of neurons recorded in the present study show correlated responses similar to those in the latter two studies, it is unlikely that pooling will obviate the necessity of using temporal integration to increase the overall signal/noise ratio.

The ROC analysis points to the importance of temporal integration for obtaining a sufficiently large number of spikes to differentiate reliably between stimuli. The results obtained with masked 40-ms target duration (Fig. 3B), a condition in which, behaviorally, the targets were discriminated as accurately as in the unmasked 20-ms condition (Fig. 1), agree with this view. Indeed, with a bin width of 80 ms, the difference between the best and worst target averaged 1.8 spikes in the 40-ms masked condition (Fig. 3B), a difference similar to the 20-ms unmasked condition (mean = 1.5 spikes) (Fig. 3A). This also suggests that the integration time for shape discrimination need not be longer than 80 ms, since the differences in number of spikes for the best and worst conditions of the 20-ms unmasked and the 40-ms masked conditions were already similar for this relatively short bin width. Thus, late phases of the response, beyond 80 ms, seem not to contribute to shape discrimination, a conclusion in agreement with other studies (37)

Both the ROC analysis of the single units as well as the above described analysis at "population level" suggest that the central nervous system integrates temporally in order to obtain a sufficient number of spikes for reliable shape discrimination. Discrimination impairment will occur if the shape-selective response is interrupted within this critical time window. As we have shown, the latter is exactly the effect of the mask in the condition of backward masking. This temporal integration could be implemented as a temporal low-pass filtering of the output of these IT units. Such low-pass filtering can only reduce the already unreliable, transient responses to the target in the masked condition. Temporal integration may already occur to some degree in some IT neurons, since we have observed 3 units (5%) for which the initial responses in the masked condition were significantly less than in the unmasked condition. One of these neurons, illustrated in Fig. 5, even failed to respond to the target in the 20-ms masked condition. Neurons of this sort may form a direct neuronal correlate of the perceptual phenomenon of backward masking.

## CONCLUSION

A patterned mask interrupts the responses to other previously briefly shown patterns, thus limiting further processing to the initial phases of the response. The present analysis of these neurophysiological data indicate that pattern backward mask-



FIG. 5. Responses of a single IT neuron to shapes presented for 20 ms followed (M) or not followed (UM) by the mask. PSTHs are unnormalized, showing the spike frequency (20-ms bin width). For other conventions, see Fig. 2.

ing occurs under those conditions in which the system needs temporal integration of the neuronal response to obtain reliable behavioral performance and in which this required integration time is cut short by the mask.

We thank P. Kayenbergh, G. Meulemans, G. Vanparrijs, and M. De Paep for technical support; Dr. W. Spileers for the search coil surgery; and Drs. B. De Bruyn, K. Lauwers, S. Raiguel, and A. A. Schoups for critical reading of the manuscript. This research was supported by the Human Capital and Mobility Program (CHRX-CT93-0267; E.E.C.) and by Grant IUAP-22 "Vision and Memory" (Belgium). R.V. is a research associate of the Belgian National Research Council (NFWO). G.K. was supported by a grant from the Belgian Science Policy Office while on leave from the Department of Physiology of the Albert Szent-Györgyi Medical University, Szeged, Hungary.

- 1. Werner, H. (1940) Am. J. Psychol. 53, 418-422.
- Breitmeyer, B. G. (1984) Visual Masking: An Integrative Approach (Oxford Univ. Press, New York).
- 3. Dennett, D. C. (1991) Consciousness Explained (Little, Brown, Boston).

- 4. Breitmeyer, B. & Ganz, L. (1976) Psychol. Rev. 83, 1-36.
- 5. Michaels, C. F. & Turvey, M. T. (1979) Psychol. Res. 41, 2-61.
- 6. Bergen, J. R. & Julesz, B. (1983) Nature (London) 303, 696-698.
- 7. He, Z. J. & Nakayama, K. (1993) Vision Res. 34, 151-162.
- 8. Nothdurft, H. C. (1993) Vision Res. 33, 1937-1958.
- 9. Felsten, G. & Wasserman, G. S. (1980) Psychol. Bull. 88, 329-353.
- 10. Gruesser, O. J., Petersen, A. & Sasowski, R. (1965) *Pflügers Arch. Gesamte Physiol.* 283, R50-51.
- 11. Schiller, P. H. (1968) Vision Res. 8, 855-866.
- 12. Bridgeman, B. (1975) Vision Res. 15, 91-99.
- Smith, E. E., Haviland, S. E., Reder, L. M., Brownell, H. & Adams, N. (1976) J. Exp. Psychol. Hum. Percept. Perform. 2, 151-161.
- 14. Weisstein, N., Harris, C. S., Bernbaum, K., Tangney, J. & Williams, A. (1977) Vision Res. 17, 341-350.
- 15. Schiller, P. H. (1965) J. Exp. Psychol. 69, 193-199.
- Desimone, R., Albright, T. D., Gross, C. G. & Bruce, C. (1984) J. Neurosci. 4, 2051–2062.
- 17. Gross, C. G., Rocha-Miranda, C. E. & Bender, D. B. (1972) J. Neurophysiol. 35, 96-111.
- Tanaka, K., Saito, H., Fukada, Y. & Moriya, M. (1991) J. Neurophysiol. 66, 170-189.
- 19. Desimone, R., Fleming, J. & Gross, C. G. (1980) *Brain Res.* 184, 41–55.
- 20. Young, M. P. (1992) Nature (London) 358, 152-154.
- Fehmi, L. G., Adkins, J. W. & Lindsley, D. B. (1969) *Exp. Brain Res.* 7, 299–316.
- 22. Seltzer, B. & Pandya, D. N. (1978) Brain Res. 149, 1-24.
- Kovács, Gy., Vogels, R. & Orban, G. A. (1995) J. Neurosci. 15, 1984–1997.
- 24. Sáry, G., Vogels, R. & Orban, G. A. (1993) Science 260, 995-997.
- 25. Judge, S. J., Richmond, B. J. & Chu, F. C. (1980) Vision Res. 20, 535–538.
- 26. Vogels, R. & Orban, G. A. (1991) J. Neurosci. 10, 3543-3558.
- 27. Kahneman, D. (1968) Psychol. Bull. 70, 404-425.
- 28. Rolls, E. T. & Tovée, M. J. (1994) Proc. R. Soc. London B 257, 9-15.
- Rolls, E. T., Tovée, M. J., Purcell, D. G., Stewart, A. L. & Azzopardi, P. (1994) *Exp. Brain Res.* 101, 473–484.
- Kovács, Gy., Vogels, R. & Orban, G. A. (1994) Soc. Neurosci. Abstr. 20, 1054.
- Cohn, T. E., Green, D. G. & Tanner, W. P. (1975) J. Gen. Physiol. 66, 583–616.
- Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. (1992) J. Neurosci. 12, 4745–4765.
- 33. Zohary, E., Hillman, P. & Hochstein, S. (1990) Biol. Cybern. 62, 475-486.
- 34. Vogels, R. (1990) Biol. Cybern. 64, 25-31.
- 35. Zohary, E., Shadlen, M. N. & Newsome, W. T. (1994) Nature (London) 370, 140-143.
- 36. Gawne, T. J. & Richmond, B. J. (1993) J. Neurosci. 13, 2758-2771.
- Tovée, M. J., Rolls, E. T., Treves, A. & Bellis, R. P. (1993) J. Neurophysiol. 70, 640-654.