Suppression and Contrast Normalization in Motion Processing

Christian Quaia,¹ Lance M. Optican,¹ and Bruce G. Cumming¹

Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, Maryland 20892

Sensory neurons are activated by a range of stimuli to which they are said to be tuned. Usually, they are also suppressed by another set of stimuli that have little effect when presented in isolation. The interactions between preferred and suppressive stimuli are often quite complex and vary across neurons, even within a single area, making it difficult to infer their collective effect on behavioral responses mediated by activity across populations of neurons. Here, we investigated this issue by measuring, in human subjects (three males), the suppressive effect of static masks on the ocular following responses induced by moving stimuli. We found a wide range of effects, which depend in a nonlinear and nonseparable manner on the spatial frequency, contrast, and spatial location of both stimulus and mask. Under some conditions, the presence of the mask can be seen as scaling the contrast of the driving stimulus. Under other conditions, the effect is more complex, involving also a direct scaling of the behavioral response. All of this complexity at the behavioral level can be captured by a simple model in which stimulus and mask interact nonlinearly at two stages, one monocular and one binocular. The nature of the interactions is compatible with those observed at the level of single neurons in primates, usually broadly described as divisive normalization, without having to invoke any scaling mechanism.

Key words: masking; motion; normalization

Significance Statement

The response of sensory neurons to their preferred stimulus is often modulated by stimuli that are not effective when presented alone. Individual neurons can exhibit multiple modulatory effects, with considerable variability across neurons even in a single area. Such diversity has made it difficult to infer the impact of these modulatory mechanisms on behavioral responses. Here, we report the effects of a stationary mask on the reflexive eye movements induced by a moving stimulus. A model with two stages, each incorporating a divisive modulatory mechanism, reproduces our experimental results and suggests that qualitative variability of masking effects in cortical neurons might arise from differences in the extent to which such effects are inherited from earlier stages.

Introduction

The response of sensory neurons to their preferred stimulus is often influenced (typically suppressed) by the presence of another stimulus ("mask"), which by itself does not drive the cell. In the visual system, these phenomena are divided into overlay suppression (also called cross-orientation suppression), observed when stimulus and mask are superimposed, and surround suppression, observed when stimulus and mask are at different spatial locations. In addition, the output of many visual neurons is nonlinearly related to the contrast of their preferred stimulus. This variable contrast gain is usually ascribed to a mechanism

called contrast normalization, which reduces the response of the cell to high-contrast stimuli and might be at least partially responsible for overlay suppression as well (Heeger, 1992). Notably, both mask suppression and contrast normalization are stronger along the dorsal visual stream, which includes the retina, the lateral geniculate nucleus (LGN), primary visual cortex (V1), the middle temporal (MT) cortex, and the middle superior temporal (MST) cortex, and particularly in neurons that are part of the magnocellular pathway that processes fast-moving stimuli (Derrington and Lennie, 1984; Kaplan and Shapley, 1986; Sclar et al., 1990; Benardete et al., 1992; Kremers et al., 2001; Solomon et al., 2002, 2006; Webb et al., 2002; Alitto and Usrey, 2008; Camp et al., 2009; Alitto et al., 2011).

Because these nonlinear suppressive mechanisms act over multiple levels of processing, it is often difficult to distinguish effects that emerge anew at one stage from those that are inherited from previous stages. Not surprisingly, such a cascade of nonlinearities and the convergence of inputs associated with the increasing size of receptive fields (RFs) along the cortical hierarchy makes the response of a cortical visual neuron a highly nonlinear function of the visual inputs that it receives indirectly and predictably only for the simplest of stimuli (David and Gallant, 2005;

Received June 6, 2017; revised Aug. 11, 2017; accepted Aug. 18, 2017.

Author contributions: C.Q., L.M.O., and B.G.C. designed research; C.Q. performed research; C.Q. contributed unpublished reagents/analytic tools; C.Q. analyzed data; C.Q., L.M.O., and B.G.C. wrote the paper.

This work was supported by the Intramural Research Program of the National Eye Institute, National Institutes of Health. We thank Boris Shella for assistance in running the experiments and two anonymous reviewers for their comments and suggestions.

The authors declare no competing financial interests.

Correspondence should be addressed to Christian Quaia, Laboratory of Sensorimotor Research, National Eye Institute, NIH, DHHS, 49 Convent Dr., Rm. 2A50, Bethesda, MD 20892. E-mail: quaia@nei.nih.gov.

DOI:10.1523/JNEUROSCI.1572-17.2017

Copyright © 2017 the authors 0270-6474/17/3711051-16\$15.00/0

Willmore et al., 2010; Nishimoto and Gallant, 2011). Accordingly, determining how these different suppression mechanisms affect a behavioral response shaped by the cooperation of populations of neurons distributed across multiple areas is quite difficult.

We explored this question by measuring, in human subjects, reflexive eye movements known as ocular following responses (OFRs), a behavioral response supported primarily by the dorsal visual pathway (Miles, 1997, 1998; Kawano, 1999; Masson, 2004; Miles et al., 2004; Takemura et al., 2007; Masson and Perrinet, 2012). Here, we focused on the effect that a mask, which does not by itself induce OFRs, has when presented simultaneously with a stimulus that in isolation produces robust OFRs. Because of their graded response to suprathreshold stimuli, OFRs are ideally suited to determine the impact of a wide range of mask contrasts on a wide range of stimulus contrasts. We mostly used static masks, simulating conditions in which a moving object is seen in front and/or behind static textured backgrounds, a common situation in real life. To further approximate the typical visual experience of foveate animals (Geisler et al., 2007), stimulus presentations were short (160 ms).

We found that a static mask can profoundly suppress the OFRs induced by a moving stimulus. The magnitude of this suppression depends in a nonlinear and nonseparable manner on the contrast and spatial frequency (SF) of both the moving stimulus and the mask and on whether mask and stimulus are superimposed or occupy different spatial locations. The range of complex interactions that we describe is captured by a cascade model in which both contrast normalization and mask-induced suppression act in a divisive (i.e., nonlinear) manner at two stages, one monocular (mainly subcortical) and one binocular (cortical). Many properties of these inferred mechanisms are commensurate with known physiology. A very simple model is thus capable of capturing the overall effect of multiple nonlinear mechanisms acting across large populations of neurons and over multiple areas.

Materials and Methods

Subjects. Three subjects (all males, aged 22–48) participated in the study. One was an author; the others had been subjects in previous ocular following experiments, but were unaware of the questions being investigated. Two additional subjects (also males, aged 32 and 55) participated in only one experiment. All subjects had normal or corrected-to-normal visual acuity and normal stereoacuity. Experimental protocols were approved by the Institutional Review Board concerned using human subjects. The study was performed in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and written informed consent was obtained from all subjects. All personal identifiable information was handled in accordance with National Institutes of Health privacy policies and regulations.

Apparatus. The apparatus used to acquire the data presented here has been described in detail previously (Quaia et al., 2016) and will only be described briefly here.

Each subject sat in a dark room and had his head stabilized using chin- and forehead-padded supports and a headband. The subject was surrounded by three CRT monitors, one in front and one on each side. The latter two were part of a Wheatstone mirror stereoscope with removable mirrors used for dichoptic presentation of stimuli. The CRT monitor facing the subject was located 525 mm from the corneal vertex, covered 41° (H) by 31° (V) of visual angle, and was set at a resolution of 1024 columns by 768 rows, and a refresh rate of 150 Hz. Only the red channel was used, since red phosphors had the shortest persistence (800 μ s rise time, 3.2 ms fall time measured using a photodiode connected to a memory digital oscilloscope), guaranteeing the absence of motion streaks on the monitor (DeAngelis and Newsome, 2004). The two side CRT monitors were arranged so that the optical distance of the monitors and the

apparent distance of the binocular image seen through the mirrors were identical (521 mm). Each monitor screen covered 50° (horizontal) by 32° (vertical) of visual angle, was set at a resolution of 1280 columns by 800 rows, and had a refresh rate of 140 Hz. Again, only the red channel was used to minimize persistence (1 ms rise time, 4 ms fall time). The refresh timing of the two monitors was tightly synchronized, with the left eye image consistently preceding the right eye image by <50 μ s. Luminance linearization was performed by interpolation following dense luminance sampling (using a Konica Minolta LS100 luminance meter) independently for each monitor.

Horizontal and vertical positions of the dominant eye were recorded using an electromagnetic induction technique (Robinson, 1963). A scleral search coil embedded in a silastin ring (Skalar) (Collewijn et al., 1975) was placed in the eye after application of topical anesthetic (proparacaine HCl). The coil output (sampled at 1000 Hz) was calibrated at the beginning of each recording session by having the subject look at targets of known eccentricity. Peak-to-peak noise levels indicated an uncertainty in eye position recording of <0.03°.

The experiment was controlled by two computers, one running the Real-time EXperimentation (REX) software package (Hays et al., 1982) to manage the work flow and to acquire and store the data, and the other connected directly to the CRT displays to generate the required visual stimuli in response to REX commands. This was accomplished using the Psychophysics Toolbox 3.0.8, a set of Matlab (The MathWorks) scripts and functions (Brainard, 1997).

Behavioral paradigm. Trials were presented in blocks and each block contained one trial for each stimulus condition. All conditions within a block were interleaved randomly. Each trial began with the appearance of a central fixation cross on a blank, midluminance (6.0 cd/m²) background. The subject was instructed to look at the center of the cross and avoid making saccadic eye movements. After the subject maintained fixation within a small (1° on the side) invisible window around the fixation point for 800–1100 ms, the fixation cross disappeared and the visual stimulus sequence (24 frames) was presented. Subsequently, the screen was blanked (again at 6.0 cd/m²), signaling the end of the trial. After a short intertrial interval, a new trial was started. If the subject blinked or if saccades were detected during the stimulus presentation epoch, the trial was discarded and repeated within the block.

Visual stimuli. All the stimuli used in the experiments described here had a mean luminance of 6.0 cd/m². Unless otherwise noted, they were presented within a circular aperture (28° diameter) centered on the screen. Outside of the aperture, the screen was blank with a luminance of 6.0 cd/m². The basic stimulus was obtained by summing the contrasts (i.e., deviations from the mean luminance) of two horizontal sinusoidal gratings, the SF and contrast of which could be varied independently. One of the gratings drifted upward or downward at 18.75 Hz; the other (the mask) was static. The initial phase of each grating was independently randomized to one of eight values (between 0° and 315° in 45° steps for the drifting grating and between 22.5° and 337.5° in 45° steps for the mask; the two were treated differently to guarantee that, even if the gratings had the same SF and contrast, there would be no frame in which the two gratings would cancel out completely). A space–time plot of unmasked and masked stimuli is shown in Figure 1. Because of the short stimulus duration, the temporal frequency (TF) spectrum of the static grating peaks at 0 Hz, but spreads out significantly to both positive and negative TFs (i.e., both directions of motion orthogonal to the orientation of the stimulus), with a full width at half-height (FWHH) bandwidth of the amplitude spectrum of 14 Hz. Therefore, a detailed characterization of the temporal properties of masking was not attempted. In most experiments, the compound stimulus was presented on a single monitor and viewed binocularly, but in some experiments, the two gratings were presented separately to each eye (dichoptic presentation).

In the last experiment a drifting horizontal sinusoidal grating (0.25 cpd, 8% Michelson contrast) was masked by a horizontal 1D random line stimulus (RLS); the mask was not limited to being static, but always had equal motion energy in both directions. The RLS was obtained by first randomly assigning either a high or a low luminance value to consecutive pairs of rows of pixels (0.08°); a high-pass filter was then applied in the Fourier domain (the gain of the filter was zero below 0.75 cpd and one

above 1.5 cpd; the transition followed a raised-cosine function), and finally the root mean square (RMS) contrast of the filtered RLS was set to 24%. Subsequent frames for the mask were computed by applying a constant phase shift to all Fourier components, but with a randomly assigned direction (constant throughout a trial) for each component. An inverse Fourier transform was then used to obtain the frame to be displayed. The resulting stimulus is broadband in SF, narrowband in orientation and in TF (within the limits imposed by our short presentation), but balanced across directions so that it generates no OFRs. We consider it superior to a classic counterphase-modulated sinewave stimulus for this study because its RMS contrast is constant throughout the presentation and thus should not modulate fast-acting contrast normalization mechanisms. Note that simply summing two RLS stimuli, each with half the contrast, drifting at the same TF in opposite directions would not have achieved the desired result: RMS contrast is a nonlinear operator and the contrast of the sum of two stimuli is equal to the sum of their contrasts only if they are identical; otherwise, it is smaller (in the limit $RMS(A) = RMS(-A)$, but $RMS(A + (-A)) = 0$). In one condition, each frame of the high-pass-filtered RLS stimulus was generated anew (temporally uncorrelated or flicker mask). Note that the phase-shifting noise used for the other masks is free of any temporal aliasing and is thus flicker free.

In our experiments, stimulus and mask had always the same orientation. It is common in neurophysiological studies to vary the orientation of the mask; for overlay conditions, orthogonal masks are most often used (hence the use of the term cross-orientation suppression). However, when using eye movements (or perception, for that matter) as a probe of suppression processes, this is problematic: in addition to suppressive effects, masks that are not parallel to the stimulus generate pattern motion signals, which contaminate the results (Quaia et al., 2016). For example, an orthogonal mask can either increase or decrease OFRs, depending on the relative contrast of mask and stimulus (our unpublished observations).

We used horizontal stimuli (and thus vertical motion) for a purely technical reason: because of slew rate limitations in the video amplifier (Pelli and Zhang, 1991), luminance distortions in CRT monitors are much lower across lines (i.e., in the vertical direction) than along lines (i.e., in the horizontal direction) and it is thus preferable to minimize luminance variations along a line (i.e., use horizontal gratings). This is especially important with high-SF noise stimuli, such as the RLS used in the last experiment, in which large luminance variations across distances of only a few pixels are possible (with vertically oriented stimuli, both mean luminance and contrast would be significantly different from the intended values).

Experimental design and statistical analysis. All of the measures reported here are based on the velocity of the instrumented eye. The calibrated eye position traces (see “Apparatus” section) were differentiated using a 21-point FIR acausal filter (47 Hz cutoff frequency). Trials with saccadic intrusions and unstable fixation that went undetected at run time were removed using an automatic procedure aimed at detecting outliers (Quaia et al., 2013). Average temporal profiles time locked to stimulus onset were then computed over the remaining trials separately for each stimulus condition.

To remove the effect of components of the eye response related to the disengagement of fixation (Boström and Warzecha, 2010; Quaia et al., 2012), we report not the raw OFR, but rather the difference between the OFRs to upward and downward motion directions. The traces and measurements reported here are thus based on the difference between the average responses to stimuli containing motion energy in opposite directions.

For each subject and experiment, data were collected until a criterion based on the signal-to-noise ratio of the responses was reached. More precisely, after each session, we extracted the data obtained in the conditions with no mask present, computed the mean trace for each of these conditions, and took the ratio between the largest eye velocity across conditions and the largest eye speed deviation from zero in the first 70 ms after stimulus onset (i.e., the period before the onset of the OFR). When this ratio became larger than 8, we considered the data collected sufficient and proceeded with the analysis of the dataset. This criterion is based

mostly on our experience of what is required to be able to estimate reliably the latency of the OFRs (which is used to determine the window over which to compute the magnitude of the OFR; see Fig. 1). For each experiment, the number of trials and sessions recorded thus varied considerably across subjects. For example, the data reported in Figure 3 required two sessions in S1, but four in S2 and S3. Several factors contribute to this variability. The noise level in the data is determined mostly by the ability of each subject to fixate precisely. This varies considerably across subjects and improves with experience, but even within each subject, there is considerable variability related to many possible environmental factors (e.g., fatigue). The signal level (the size of each subject's OFRs) also varies widely across subjects. Because all of these factors are outside of the experimenter's control, the number of trials/sessions cannot be established a priori.

Between two and four subjects participated in each of the main experiments. A few control experiments aimed at ruling out some unlikely interpretations of the data were performed in only one subject. The cascade model (see “Model fitting” section) was applied to the data collected from multiple subjects in the main experiments. In addition, it was applied to a set of experiments (presented in Fig. 13) that were performed in only one subject. There are several reasons that we chose to use just one subject in this last experiment. First, this experiment was used to refine and further strengthen conclusions that had been reached based on data collected from multiple subjects in the other experiments. Second, the conditions tested can be seen as an interpolation within the set of conditions tested in the main experiments (the same parameters were varied within the same range of values). Third, in the main experiments, there were no qualitative differences across subjects. Finally, this final experiment required collecting almost 32,000 trials from the subject with the highest signal-to-noise ratio (the criterion used to determine which subject to use in this experiment); from other subjects, up to 60,000 trials might have been necessary to obtain data of sufficient quality to constrain the model meaningfully, which would have required up to 4 months of recordings.

All statistical analyses, including computations of standard errors and significance values, were performed using nonparametric, bootstrap-based methods (Efron, 1982). A detailed description of the bootstrap procedures used has been described previously (Quaia et al., 2012).

Model fitting. To quantify the extent of suppression as a function of mask SF, we fit the normalized response (i.e., the ratio between the response to stimulus plus mask and the response to the stimulus alone) using a logarithmic Gaussian function as follows:

$$\mathcal{N}(SF_m | SF_d, c_d, c_m) = 1 - G e^{-\left(\frac{2 \sqrt{\log(2)} \log_2(SF_m/\mu)}{\gamma}\right)^2} \quad (1)$$

where SF_m is the SF of the mask. The SF of the drifting stimulus (SF_d), its contrast (c_d), and the contrast of the mask (c_m) are here shown as given constants, indicating that the fit parameters G , μ and γ may be a function of those stimulus parameters. This formulation is slightly different from the commonly used one, in that the logarithm is in base 2, and the dispersion parameter has been scaled so that γ equals the FWHH bandwidth of the function in octaves (a measure commonly reported in neurophysiological experiments). We further defined as low and high cutoff frequencies the values of SF_m for which the normalized response equals 0.8 (i.e., the mask induces a 20% attenuation):

$$SF_{L/H} = \mu 2^{\pm \gamma \log(5G)/(2 \sqrt{\log(2)})} \quad (2)$$

To quantify how both stimulus and mask contrast contribute to determining the magnitude of the OFR, we used the Naka–Rushton function (also known as the Michaelis–Menten equation). The basic formulation relies on three parameters to quantify the response to the drifting stimulus as a function of its contrast:

$$\mathcal{R}(c_d | SF_d) = A \frac{c_d^n}{c_d^n + c_{50}^n}$$

where A is a gain factor (which for the OFRs takes into account also the sensorimotor gain and thus varies widely across subjects), n determines

the slope of the curve, and c_{50} is the semisaturation contrast (i.e., the contrast at which the response drops to $A/2$) and places the curve along the c_d axis. This equation quantifies the overall contrast gain of the system, $R(c_d)/c_d$, and thus accounts for contrast normalization processes (sometimes referred to as tuned inhibition).

The effect of the mask can be introduced by having the parameters of the function vary with mask contrast c_m . In some cases, the effect is well described by a change in the parameter c_{50} alone. In this case, it is said that the mask acts in a divisive manner and thus we call this a divisive model:

$$\mathcal{R}(c_d, c_m | \text{SF}_d, \text{SF}_m) = A \frac{c_d^n}{c_d^n + c_{50}(c_m)^n} \quad (3)$$

By fitting our data with a series of Naka–Rushton functions in which the value of c_{50} was optimized independently for each value of the mask contrast (while keeping A and n constant), we discovered that, in the context of our experiments, the relationship between c_{50} and the contrast of the mask c_m is well captured by a quadratic function in log–log space as follows:

$$c_{50}(c_m) = 10^{\sigma_2 c_m^2 + \sigma_1 c_m + \sigma_0} \quad (4)$$

where

$$c_{ml} = \log_{10}(\max(c_m, 1.0))$$

This latter formula, which limits c_{ml} to positive values, assumes that mask contrasts below 1% (which were never used in our experiments) are not effective and 10^{σ_0} is thus the value of c_{50} appropriate to fit the contrast response curve in the absence of a mask.

The mask might also act by changing the gain factor A , thus scaling the response. The resulting scaling model would then be:

$$\mathcal{R}(c_d, c_m | \text{SF}_d, \text{SF}_m) = A(c_m) \frac{c_d^n}{c_d^n + c_{50}^n}$$

We found that, in all of our experiments, a divisive component always contributed to the suppressive effects, but in some cases, a scaling component was also necessary to account for the data. We thus defined a mixed model incorporating divisive and scaling terms as follows:

$$\mathcal{R}(c_d, c_m | \text{SF}_d, \text{SF}_m) = A(c_m) \frac{c_d^n}{c_d^n + c_{50}(c_m)^n} \quad (5)$$

Empirically, we found that the relationship between A and the contrast of the mask c_m can be captured by a quadratic function in log space as follows:

$$A(c_m) = A_2 c_{ml}^2 + A_1 c_{ml} + A_0 \quad (6)$$

where A_0 is the value of A appropriate to fit the contrast response curve in the absence of a mask.

Function fitting was performed using a simplex optimization algorithm in Python (using the numpy library). We did not minimize the summed squared error, but rather the χ^2 measure, as follows:

$$\chi^2 = \sum_i \left(\frac{y_i - \hat{y}_i}{s_i} \right)^2$$

where, for each experimental condition i , we indicate with y_i the mean OFR across trials, s_i the standard error of y_i , and $\hat{y}_i$ the value of the fitted function for that condition. More consistent responses are thus weighted more heavily in determining the quality of a given fit, maximizing the likelihood of the model given the data. To minimize the risk that the optimization algorithm might find a suboptimal solution (local minimum in the χ^2 landscape), we repeated the optimization procedure multiple times initializing the search with different initial conditions, selected using a Sobol algorithm.

When multiple fitting functions are available, a process of model selection must be used to determine which one best accounts for the data. There is not a single or best way to determine which function is the most

appropriate for a given dataset. Here, we report two measures that quantify the goodness of fit. The first one is χ^2 , which describes how well the function fits the data given the variability in the measurements, with lower values indicating a better fit. This measure does not account for the number of parameters of the fitting function and because it is to be expected that functions with more parameters will fit the data better (at least as long as the functions are nested), it cannot be used to determine which function is best: a way to penalize the fit based on the number of parameters must be used. We report the small-sample Akaike's information criterion (AICc) (Burnham and Anderson, 2002). Although better fits are associated with lower values of AICc, its absolute value is not meaningful; instead, the difference between the AICc computed for the different models is reported. It is generally accepted that a difference of at least 10 is required to conclude with confidence that a function is more appropriate than another one (Burnham and Anderson, 2002). If adding parameters to a function does not reduce AICc by at least 10 points, then we consider the function with fewer parameters as the “best”-fitting function. If multiple functions having the same number of parameters have AICc values that are not different by at least 10 points, we consider them as equivalent. Here, we report ΔAICc , the difference between the AICc for the divisive plus scaling model and that for the divisive model. Unless this value is smaller than -10 , we conclude that the divisive model is sufficient to account for the data. It is important to realize that, with all model selection procedures, a lower signal-to-noise ratio in the data (which can be due to large variability in the data or to a small number of trials over which to average the data) biases the result toward selecting the model with fewer parameters. A more complex model might thus be rejected only because there is not enough data or the data are too noisy to be able to demonstrate a clear improvement in fit quality. Acquiring enough data of good enough quality is thus essential to apply a model selection procedure meaningfully.

Temporal filter analysis. In the last experiment, we used a nonstatic mask: an RLS in which the motion energy was concentrated around a single TF, but was split equally in both directions (see “Visual stimuli” section). From the OFRs, we estimated the strength of the mask as a function of its center TF as follows:

$$\mathcal{M}(\text{TF}_m) = 1 - \mathcal{N}(\text{TF}_m)$$

where N is the normalized response. Using a linear relationship to derive masking strength from N is justified by the fact that, when the contrast of the drifting grating is low (8% in our case) and $N > 0.5$, the relationship between N and the contrast of the mask in the other experiments reported here is well approximated by a linear function. To simulate the response of nondirectional spatiotemporal filters to our masking stimuli (160 ms presentations, at 150 Hz), we then generated 1000 random line stimuli spanning the range of TFs covered by the masks in our experiment and passed them through a quadrature pair of spatial filters (odd and even Gabor functions, with a 1 cpd center SF and a FWHH bandwidth of two octaves, typical of cells in area V1). Next, we convolved the output of the spatial filters with one of two temporal filters. The first one was monophasic, representing a low-pass filter:

$$LF(t) = \left(\frac{t}{\tau} \right)^n e^{-t/\tau}$$

The second was instead biphasic, representing a band-pass filter (Adelson and Bergen, 1985):

$$LF(t) = (kt)^n e^{-kt} \left(\frac{1}{n!} - \beta \frac{(kt)^2}{(n+2)!} \right)$$

Finally, we computed the mean response in a 70 ms time window, the window that we used to quantify the OFRs in our experiments. To a first-order approximation, this model simulates the combined activity of complex cells that are not direction selective and complex cells that are direction selective, but regardless of their directional preference. The parameters of the temporal functions (two for the monophasic function, three for the biphasic function) were varied to best fit the data (using a simplex optimization algorithm, as described in the previous section).

Figure 1. Masking of sample OFRs. The sudden onset of a drifting horizontal sinusoidal grating (0.25 cpd SF, 18.75 Hz TF, 32% contrast) induces a short-latency (~ 70 ms in humans) vertical OFR (black). When a static horizontal sinusoidal grating (0.5 cpd SF, 32% contrast) is added to the stimulus, the ensuing OFR is much smaller (gray), although its latency is unchanged. Most tuning properties of the OFR are highly similar across subjects, although the absolute magnitude of a response to a given stimulus can vary widely across subjects (vertical scale corresponds to 0.25 deg/s). We quantify the strength of the response by computing the average eye speed in a time window that goes from the onset of the response to twice the latency (open-loop period), indicated here with a gray bar above the abscissa. Space–time diagrams show the temporal evolution of unmasked and masked stimuli. Data are shown from three subjects (S1, S2, and S3).

Because of their short duration, both the moving stimulus and the mask are not particularly narrowband in TF and thus are not well suited to study the temporal properties of either the ocular following system or of the masking mechanism. Unsurprisingly, we thus found that both a monophasic and a biphasic temporal function could capture the data quite well. However, the goal of this experiment was to use the filters to predict the masking strength of a flickering masking stimulus, which was not used in the fitting procedure.

Results

Superimposed static mask: dependence on SF and contrast

We started our investigation on the impact of a mask on the processing of motion information by measuring the vertical OFRs induced by a drifting (18.75 Hz) horizontal sinusoidal grating (stimulus) alone or in the presence of a superimposed static horizontal grating (mask). Both gratings had the same Michelson contrast (32%). The two gratings appeared together suddenly on a mid-luminance background and the stimulus started drifting immediately (phase change of 45° in each successive frame). The entire presentation lasted 24 frames (160 ms).

In Figure 1, we show the time course of two sample OFRs for three subjects. One trace (black) shows the OFR generated when the drifting stimulus (0.25 cpd, 32% contrast) was presented in isolation and the other (gray) shows the effect of adding to this drifting stimulus a static mask (0.5 cpd) having the same contrast. Space–time plots of both stimuli are shown in the leftmost panel. In all subjects, the mask strongly suppressed the response. To quantify this suppressive effect, we computed the average eye speed in a time window (slightly different for the three subjects and indicated with a horizontal bar above the time axis in Fig. 1), which is contained within the so-called open-loop period (the period during which the motion of the eyes cannot affect the processing of the visual stimulus). From here on, we focus on this measure, which we simply call the OFR.

To estimate the SF tuning of the suppressive mechanism(s), we varied independently the SF of stimulus and mask while keeping the contrast fixed at 32% for both stimulus and mask. In Figure 2A we show how the normalized response N (defined as the ratio between the OFR measured when the mask is present and that induced by the drifting grating alone) varies as a function of the SF of the two gratings. The drifting grating had one of three SFs: 0.125 cpd (red), 0.25 cpd (blue), or 0.5 cpd (green); the

mask had one of six SFs (from 0.0625 to 2 cpd in one-octave steps). We fit the normalized response as a function of the SF of the mask separately for each SF of the drifting grating, using a log-Gaussian function (see Materials and Methods, Eq. 1). The SF associated with the strongest suppression (μ in Eq. 1, indicated with color-matched dots and SEM bars at the top of each panel) shifts significantly as the SF of the stimulus is varied. Parameter values for the fits are listed in Table 1.

In Figure 2 we also show, for each subject, mean and SEM of μ (Fig. 2B), bandwidth γ (Fig. 2C), low (Fig. 2D), and high (Fig. 2E) SF cutoff (SF at which the normalized response equals 0.8, Eq. 2) of the fitting functions, as a function of the SF of the moving grating. The low SF cutoff (Fig. 2D) shifts approximately in step with the SF of the moving grating and is approximately two octaves lower (1.78 octaves, average across all subjects and conditions), indicating that masks at a larger scale than the stimulus are not very effective. The high SF cutoff (Fig. 2E) is instead fairly constant at ~ 5 cpd (4.76 cpd, average across all subjects and conditions). It follows that the bandwidth (Fig. 2C) decreases as the SF of the moving grating increases. Obviously, neither the absolute SF of the mask nor the relative SF of the two gratings fully determine the extent of suppression (Fig. 2B). Because the strongest effects are not seen when stimulus and mask have the same SF, they cannot simply be imputed to the impact that the mask has on the response of the linear filter that detects the stimulus. This is especially true for static masks, which are poor stimuli for direction-selective cells. This linear interaction, which is the subject of many psychophysical studies of masking, might play some role, but the bulk of the effects observed here reflect nonlinear interactions either in the feedforward pathway or through lateral or feedback inhibition.

To gain further insights into the mechanism(s) underlying suppression, we then fixed the SF of both gratings and varied their contrasts independently (between 2.5% and 40%, in one-octave steps); in addition, the mask could also be absent (0% contrast). We selected a pair of SFs (0.25 cpd for the drifting grating and 0.5 cpd for the mask) for which a strong suppression was induced by the mask in all three subjects in the first experiment. In Figure 3, we plot the OFRs as a function of the contrast of the drifting grating. In the no-mask condition (Fig. 3, black), the response saturates at high contrast, indicating that a strong contrast normalization mechanism is at work, as expected from previous reports (Miles et al., 1986; Sheliga et al., 2005; Miura et al., 2006). As the contrast of the mask is increased progressively (see color key), the response is increasingly suppressed. We fit two Naka–Rushton functions to the data (see Materials and Methods), one accounting for a divisive effect of the mask (Eq. 3) and the other allowing for both divisive and scaling effects (Eq. 5). The fits shown in Figure 3 were obtained using the divisive model: all curves have a common value for A and n , but c_{50} is a function of mask contrast (Eq. 4). The high quality of the fits indicates that the effect of the mask is simply to shift the contrast response curve for the drifting stimulus to the right by an extent that is a function of the contrast of the mask. Adding a scaling component to the model did not improve the fits significantly. The overall effect of the mask is thus compatible with any of the various divisive nor-

malization schemes proposed in the past to account for contrast gain control and suppressive effects in cortical and subcortical neurons (Heeger, 1992; Carandini et al., 1997). Fit parameters are reported in Table 2.

Based on neurophysiological studies (Wiesel and Hubel, 1966; Dreher et al., 1976; Krüger, 1977) and psychophysical studies (Breitmeyer and Williams, 1990; Breitmeyer and Breier, 1994; Edwards et al., 1996; Pammer and Lovegrove, 2001; Bedwell et al., 2003, 2006; Chapman et al., 2004), it has been proposed that red illumination might interfere with the proper functioning of the magnocellular pathway. Because of the strong evidence that OFRs are mediated by magnocellular neurons (Takemura and Kawano, 2002; Miles and Sheliga, 2010; Masson and Perrinet, 2012), this raises the possibility that our use of red stimuli might have biased our results. To examine this hypothesis, we tested one subject using a subset of the conditions used in the previous experiment (unmasked and 10% mask, corresponding to the black and green curves in Fig. 3, respectively), but using achromatic (i.e., gray) stimuli and background. All responses at all contrast levels in both masked and unmasked conditions were not significantly different from those found with red illumination (data not shown).

If the strength of suppression were a separable function of mask SF and contrast, then the two relationships just described could be used to predict the attenuation induced by the mask under a wide range of conditions. To test this hypothesis, we replicated the last experiment using a different pair of SFs for the two gratings (0.125 cpd for the drifting grating and 1.0 cpd for the mask), chosen to be much farther apart (three octaves) than those used before. The pattern of mask-induced response attenuation that we found (Fig. 4) was very different from that reported above, now clearly showing a scaling of the contrast response. Consequently, a purely divisive model like the one used above provided poor fits (data not shown). The suppressive effects were, however, well captured by the function that included both a divisive and a scaling component, each a function of the contrast of the mask (according to Eqs. 4 and 6). In Figure 4, we show the fits obtained with the divisive plus scaling model (parameter fits are listed in Table 2).

The data in Figures 3 and 4 differ in both the SF of the mask (0.5 vs 1.0 cpd) and the SF of the drifting grating (0.25 vs 0.125 cpd). It is thus unclear whether the scaling component arises because, in the second experiment, the SF of the mask is higher, the SF of the drifting grating is lower, or the SF separation between mask and stimulus is increased (three octaves instead of one). To address this question, we measured responses to two additional combinations of stimulus and mask SF. First, in two subjects (S1 and S3), we paired a 1.0 cpd mask (as in Fig. 3) with a 0.5 cpd stimulus, thus reducing the SF separation to one octave (as in Fig. 3). Second, in one subject (S1), we paired a 0.125 cpd stimulus (same as in Fig. 4) with a

Figure 2. SF tuning of masking suppression. **A**, When the SF of stimulus and mask are varied while keeping their contrast fixed (32% each), the extent of suppression varies widely. Here, three different SFs are used for the stimulus: 0.125 cpd (red), 0.25 cpd (blue), and 0.5 cpd (green). Each of these is paired with six SFs of the mask, from 0.0625 to 2.0 cpd in one-octave steps. For each SF of the stimulus, a SF of the mask can be found that exerts almost complete suppression. In all three subjects, the SF tuning of suppression varies considerably as a function of the SF of the stimulus (log-Gaussian fits are shown; Eq. 1). Small colored dots and bars indicate mean and SEM of the location of the trough in each fit. Parameters of the fitting functions (mean and SEM) as a function of the SF of the stimulus are shown in **B–E**. **B**, μ , corresponding to the strongest attenuation. **C**, FWHH bandwidth (γ). **D**, Low SF cutoff (SF_L). **E**, High SF cutoff (SF_H). Values for the parameters are also listed in Table 1.

Table 1. Fit parameters for SF tuning of suppression

Figure	Subject	SF_d (cpd)	c_d (%)	c_m (%)	G	μ (cpd)	γ (oct)
2	S1	0.125	32	32	0.92 (0.90–0.94)	0.45 (0.43–0.47)	4.12 (3.91–4.33)
		0.25			0.86 (0.84–0.88)	0.52 (0.51–0.53)	3.17 (3.04–3.30)
		0.5			0.84 (0.81–0.87)	0.73 (0.71–0.76)	2.36 (2.21–2.53)
	S2	0.125			0.89 (0.84–0.93)	0.43 (0.39–0.47)	4.09 (3.68–4.53)
		0.25			0.91 (0.88–0.94)	0.53 (0.51–0.56)	2.87 (2.66–3.10)
		0.5			0.86 (0.82–0.90)	0.78 (0.75–0.82)	2.42 (2.21–2.62)
	S3	0.125			0.98 (0.92–1.06)	0.40 (0.37–0.45)	3.35 (2.93–3.77)
		0.25			0.83 (0.78–0.88)	0.56 (0.52–0.61)	3.23 (2.80–3.68)
		0.5			0.87 (0.78–1.03)	0.87 (0.78–0.96)	1.99 (1.55–2.37)
5A	S1	0.25	20	5	0.29 (0.26–0.34)	0.80 (0.72–0.87)	2.11 (1.60–2.79)
				10	0.56 (0.53–0.59)	0.61 (0.58–0.64)	2.13 (1.90–2.36)
				20	0.86 (0.84–0.89)	0.46 (0.45–0.48)	2.83 (2.65–3.00)
				40	0.99 (0.97–1.01)	0.45 (0.44–0.47)	3.95 (3.76–4.12)
5B	S1	0.25	5	5	0.71 (0.65–0.78)	0.48 (0.44–0.52)	2.65 (2.31–3.01)
				10	0.77 (0.74–0.81)	0.50 (0.47–0.53)	2.84 (2.56–3.11)
				20	0.83 (0.80–0.86)	0.51 (0.49–0.53)	2.70 (2.50–2.90)
				40	0.85 (0.82–0.88)	0.50 (0.48–0.52)	3.06 (2.84–3.28)
7	S1	0.25	5		0.35 (0.33–0.38)	0.40 (0.37–0.43)	3.89 (3.47–4.32)
					0.40 (0.38–0.43)	0.35 (0.33–0.38)	4.38 (4.02–4.80)
	S3	0.25	5		0.21 (0.18–0.29)	0.40 (0.31–0.52)	3.47 (1.57–4.76)
					0.26 (0.20–0.36)	0.48 (0.37–0.59)	2.77 (1.62–3.89)

Parameter values are for fits to Equation 1. Mean and 68% CIs are indicated for each parameter. CIs on the parameter values were obtained using standard bootstrapping procedures. Values that do not vary within a set of conditions are only shown on the first row.

Figure 3. Contrast tuning of masking suppression. When the contrast of stimulus and mask are varied while keeping their SF fixed (0.25 cpd for the stimulus, 0.5 cpd for the mask), the extent of suppression varies as a function of both. We used five contrast levels for the stimulus, from 2.5% to 40% in one-octave increments and five for the mask (see color key), which could also be absent (0% contrast). A Naka–Rushton function in which the mask acts in a divisive manner (Eq. 3) fits the data remarkably well in all three subjects. Parameter values for the fitting function are reported in Table 2.

0.25 cpd mask (again, one octave separation). In all cases $\Delta AICc$ was positive (ranging between 6.09 and 6.29), so the purely divisive model was sufficient to account for the mask-induced suppression (data not shown). We conclude that the SF separation between mask and stimulus is the determining factor: When the SF of the mask and stimulus are similar, a purely divisive model is sufficient to account for the masking effect; when they are far apart, a scaling component emerges.

Increasing the separation between the SFs of stimulus and mask revealed a significant scaling component of the suppressive mechanism not observed with nearby SFs. This implies that the effect of the mask is not separable in SF and contrast. Reaching this conclusion based exclusively on a model selection procedure is hazardous because it depends on the particular models used, the model selection criterion, the quality of the data, and the effectiveness of the masks used (if the mask is ineffectual or completely suppresses the responses, then both models fit equally well and the divisive one is thus selected only by virtue of having fewer parameters). As a direct experimental test of non-separability, we fixed SF (0.25 cpd) and contrast (20%) of the drifting grating and varied the contrast and SF of the mask (Fig. 5A). If separability held, then the SF of the mask yielding the strongest attenuation (fit parameter μ , indicated with small color-matched dots and SEM bars at the top of each panel in Fig. 5) should always be the same. We instead found that μ increases as the contrast of the mask is reduced (parameters of the fitting functions are listed in Table 1). Even ignoring the fits, it is apparent from looking at the data that, when the contrast of the mask is 20% (cyan), a 0.5 cpd mask is more suppressive than a 1.0 cpd mask, but when the mask contrast is decreased to 5% (gray), the opposite is true. This is compatible with our previous finding that increasing the separation between the SFs of mask and stimulus engages a scaling mechanism, which is generally more effective than a divisive mechanism when the contrast of the mask is low compared with that of the stimulus. This experiment further revealed that the bandwidth of the mechanism increases (for a constant contrast of the stimulus) with the contrast of the mask. Neither the shift in peak nor the broadening of the tuning curve are predicted by a classic normalization model (see Fig. 7D in Heeger, 1992), but cells with such behavior have been observed in cat primary visual cortex (see Fig. 9C in Bonds, 1989). Note that, if the contrast of both stimulus and mask are equal, then neither the most effective mask SF nor the bandwidth of the suppressive effect vary with

contrast (Fig. 5B). The SF bandwidth of the mechanisms can thus be specified only for the relative contrast at which it has been measured. The same mechanism can appear highly tuned if the contrast of the mask, relative to that of the stimulus, is low (with an FWHH bandwidth of ~ 2 octaves), but much less so if it is high (with an FWHH bandwidth of 4 octaves).

These results are puzzling if considered in terms of a single masking mechanism and strongly suggest that two (or possibly more) nonlinear mechanisms may be at work. To better characterize these putative mechanisms, we explored the impact of manipulating other properties of the mask and stimulus, starting with their spatial overlap.

Nonoverlapping static mask:

dependence on SF and contrast

In the study of vision, suppressive influences are divided into those that originate from the same area of the visual field as the stimulus and those that originate from the surrounding region of the visual field. When studied in single RF (CRF) or from the surround. The stimuli that we used above were very large, much larger than the size of RFs in visual areas up to and including area MT, and thus probably engage both center and surround mechanisms in most visual areas. Although analysis of OFRs cannot be used to distinguish these mechanisms in a rigorous manner, reducing the size of stimulus and mask and separating them in space can provide some indication of the relative strength of center and surround mechanisms (Barthélemy et al., 2006).

We thus presented stimulus (0.25 cpd, 32% contrast) and mask (0.5 cpd, 32% contrast) within narrow vertical apertures. The mask and the stimulus could be superimposed (overlap condition) or separated in space (surround condition). We first ran an experiment in which, as is typically done, the drifting stimulus was presented within a centrally placed aperture (4° wide and 28° high) and the mask was either superimposed on the stimulus (indicated as -2° gap in Fig. 6) or composed of two apertures, 8° wide and 28° high, one on each side of the stimulus, with a variable gap between the two. The eccentricity of the inner edge of the apertures ranged from 2° (no gap) to 7.3° (5.3° gap). We found that the attenuation was maximal when stimulus and mask were superimposed and decreased rapidly as the gap increased (Fig. 6A). This finding could be interpreted as indicating that masking suppression arises mostly from the CRF and not from the surround. However, this result might also arise if the mask is less effective when presented in the periphery of the visual field. Both of these interpretations are compatible with reports that, in a center-surround configuration with the moving stimulus presented in a 20° central disk and the static mask presented in an annular surround, a static mask has no effect (Barthélemy et al., 2006). To clarify the contribution of relative and absolute mask location, in another experiment, we placed the mask aperture (4° wide, 28° high) centrally and varied the location of the stimulus, which was shown through two apertures (each 2° wide and 28° high). We also measured the OFRs induced by the stimulus alone at all locations (thus explicitly accounting for the effect of absolute stimulus location) and computed the ratio between the response to the stimulus without and with the static mask. The

Table 2. Fit parameters for contrast tuning of suppression

Figure	Subject	SFd (cpd)	SFm (cpd)	n	σ_0	σ_1	σ_2	$A(A_0)$	A_1	A_2	χ^2	ΔAIC_c		
3	S1	0.25	0.5	2.26	0.66	0.233	0.364	1.50	N/A	N/A	47.16	−0.22		
				2.45	0.66	0.211	0.323	1.51	−0.000	−0.200	40.34			
	S2			1.97	0.69	0.234	0.352	1.12	N/A	N/A	17.81	6.66		
				1.98	0.69	0.231	0.352	1.12	−0.000	−0.006	17.88			
	S3			2.96	0.59	0.408	0.194	0.81	N/A	N/A	13.47	4.53		
				3.18	0.60	0.373	0.188	0.82	−0.025	−0.057	11.40			
4	S1	0.125	1.0	1.52	1.00	0.276	0.187	1.40	N/A	N/A	100.90	−57.04		
				1.93	0.95	0.249	0.000	1.33	−0.121	−0.286	37.27			
	S3			1.45	0.83	0.403	0.153	0.82	N/A	N/A	28.54	−11.39		
				2.00	0.80	0.313	0.000	0.78	−0.084	−0.150	10.56			
	S1			0.25	0.5	1.77	0.84	0.000	0.190	1.80	N/A	N/A	243.81	−163.77
						2.20	0.86	0.000	0.088	1.95	−0.062	−0.215	73.45	

Alternating lines show parameters values for fits of the same dataset with the divisive (Eq. 3) and divisive plus scaling (Eq. 5) models of suppression. ΔAIC_c indicates the decrease in the AICc measure obtained with the divisive plus scaling model. If this value is smaller than -10 , there is strong evidence to support this model. Otherwise, the purely divisive model should be considered sufficient to account for the data. Values that do not vary within a set of conditions are only shown on the first row.

N/A, Not applicable.

Figure 4. Contrast tuning of masking suppression. Same as Figure 3, but the SFs of stimulus and mask are now farther apart (0.125 cpd for the stimulus, 1.0 cpd for the mask). In this case, a purely divisive model does not provide a good fit to the data; a model that incorporates both divisive and scaling effects (Eq. 5), the fits of which are shown, is necessary. Parameter values can be found in Table 2.

results are plotted in Figure 6B as a function of the gap between stimulus and mask. Once again, suppression is strongest when stimulus and mask are superimposed (-2° gap); however, the suppression does not fade away gradually with separation, it quickly asymptotes to an intermediate level. A surrounding stimulus can thus exert a considerable suppression on a behavioral response even when it is far from the classical RFs of subcortical and early cortical neurons.

Two different mechanisms might mediate this suppression phenomena. First, it might be the result of long-range surround suppression in area V1 (Cavanaugh et al., 2002a), so its tuning properties might very well be different from those seen with overlay suppression. Alternatively, it might result from interactions that occur in later areas, where the RFs are large enough that their CRF encompasses both stimulus and mask in our configuration. In this case, even though mask and stimulus do not overlap physically, it would qualify as overlay suppression and the tuning properties of masking would be expected to be largely the same whether stimulus and mask overlap or are (moderately) separated. To test these alternatives, we measured the tuning properties of suppression when stimulus and mask are separated in space by a 1° gap, with the mask in the center and the stimulus in the surround. Given the low SF of our stimuli, this is by all measures a small separation.

Figure 5. Nonseparable interactions between contrast and SF. **A**, When the SF and contrast of the stimulus are kept constant (0.25 cpd, 20%) and those of the mask are varied (SF on the abscissa, see color key for contrast), the mask SF resulting in the strongest suppression varies as a function of mask contrast and so does the bandwidth of the effect. **B**, When contrast of mask and stimulus are the same (see color key), the SF tuning of suppression is essentially invariant, with only a slight broadening and deepening of the curves with increasing contrast. Fit parameters can be found in Table 3.

First, we varied the SF of the mask (from 0.0625 to 2 cpd in one-octave steps) for two values of the SF of the stimulus (0.25 and 0.5 cpd). Both stimulus and mask had 32% contrast. We found (Fig. 7, Table 1) that the SF of the mask producing the strongest suppression was the same for both SFs of the stimulus. This is obviously quite different from what we found with overlapping stimuli (Fig. 2). Next, we varied the contrast of stimulus and mask independently while keeping their SF constant (0.25 and 0.5 cpd, respectively). Whereas a divisive model accounted remarkably well for the suppression observed in the overlapping condition with these same SFs (Fig. 3), it failed with nonoverlapping stimuli. A model including both a divisive and a scaling component was necessary to obtain acceptable fits to the data (Fig. 8, Table 2), similar to what we found for the overlapping conditions in which the SF of grating and mask were widely separated (Fig. 4). Therefore, a mask that exerts a strong and purely divisive suppression when it overlaps the stimulus produces a milder suppression, but one that includes a scaling component, when presented at a different location.

These results strongly argue in favor of suppression by non-overlapping stimuli being mediated by surround interactions in area V1, not by overlay mechanisms in late cortical areas (MT or

Figure 6. Suppression in center-surround configurations. **A**, When the stimulus (0.25 cpd, 32% contrast) is presented in a central aperture and the mask (0.5 cpd, 32% contrast) flanks it with a variable gap (abscissa; negative values indicate overlap), the strength of suppression decreases as the gap increases. The spatial configuration of stimulus and mask apertures is represented graphically: gray/dot represent the static mask, black/arrow represent the drifting grating; both are presented in rectangular apertures, arranged as shown. **B**, When the mask is presented in a central aperture and the stimulus flanks it with a variable gap, the strength of suppression varies only slightly with gap size, although it is still considerably weaker than in the overlap condition.

Figure 7. SF tuning of masking for nonoverlapping masks. Same format as in Figure 2. Two different SFs, 0.25 cpd (blue) and 0.5 cpd (green), are used for the stimulus, whereas the mask can have one of six SFs. Both have 32% contrast and are arranged in a center-surround configuration with a small (1°) gap. Log-Gaussian fits to the data are shown (parameter values in Table 1). When stimulus and mask overlap, the two SF tuning curves are shifted relative to each other with the one associated with the lower stimulus SF having its trough at a lower mask SF (Fig. 2). Now the two curves are very similar and the location of their trough is not significantly different. The curves are also considerably shallower because the strength of suppression is weaker in the center-surround configuration (Fig. 6).

MST). There is, however, a potential confound: in our overlap experiments, both stimulus and mask were much larger than those used in the nonoverlapping condition. We therefore measured again, in one subject (S1), the SF-tuning functions in the overlapping condition, but using the smaller stimulus sizes used in the surround experiment. When SF was varied, we found a shift in the attenuation curves as a function of the SF of the stimulus similar to that seen with large stimuli. When stimulus and mask were shown in a central aperture (4° wide, 28° high), μ was 0.50 cpd [68% confidence interval (CI): 0.48–0.52 cpd] when $SF_d = 0.25$ cpd and 0.68 cpd (68% CI: 0.66–0.70 cpd) when $SF_d = 0.5$ cpd; when they were shown in two peripheral apertures (each 2° wide, 28° high, separated by a 4° gap), μ was instead 0.46 cpd (68% CI: 0.44–0.48 cpd) when $SF_d = 0.25$ cpd and 0.62 cpd (68% CI: 0.59–0.65 cpd) when $SF_{sd} = 0.5$ cpd. Similarly, when contrast was varied, a purely divisive model was sufficient to account for the suppressive effects when mask and stimulus were both shown in the two peripheral apertures ($\Delta AICc = 6.59$).

Figure 8. Contrast tuning of masking for nonoverlapping masks. Same format and SF for stimulus (0.25 cpd) and mask (0.5 cpd) as in Figure 3. Because stimulus and mask do not overlap, a wider range of contrasts could be explored. In the overlap condition, a purely divisive model was sufficient to account for the effect of the mask (i.e., the mask mostly shifted the contrast response function to the right without changing its slope). In the center-surround condition shown here, a significant scaling component of suppression is also present (note how the slope of the curves decreases as mask contrast increases). Fit parameters are listed in Table 2.

Therefore, overlap (or lack of it) is the determining factor, not size.

Binocular properties

The effects reported above could arise at any of a number of neural processing stages or might more likely reflect a cascade of effects across stages. Teasing out the contribution of individual processing stages is quite difficult. However, it is possible to at least determine whether the suppressive effects are confined either to neurons that receive inputs from only one eye (retina, LGN, and the input layer of area V1) or to neurons that receive inputs from both eyes (the rest of cortex). If suppression is stronger when stimulus and mask are seen by the same eye (monocular condition) than when they are presented to different eyes (dichoptic condition), then it can then be argued that suppression has a subcortical component, which is then inherited by later stages.

We presented, within a 28° circular aperture, stimulus (0.25 cpd SF, 32% contrast) and mask (0.5 cpd SF, 32% contrast) either to the same eye (monocular conditions) or one to each eye (dichoptic conditions). During monocular presentation, the other eye saw a mid-luminance blank screen. We found (Fig. 9) that OFR suppression induced by the mask was significantly stronger in the monocular (orange) than in the dichoptic (blue) condition, indicating that, for these choices of mask and stimulus, a substantial fraction of the suppression originates at the subcortical level. Note that the significant suppression observed when a dichoptic mask is presented (normalized OFR significantly smaller than 1) cannot be unequivocally imputed to a cortical mechanism of masking suppression: other, less specific cortical mechanisms such as binocular rivalry might also be responsible (Quaia et al., 2016).

The large stimulus used here activates a large number of subcortical and cortical neurons and, for many of them, it probably engages both center and surround mechanisms. In an attempt to isolate the contribution of surround mechanisms, we presented the static mask (0.5 cpd) in a central aperture (4° wide, 28° high) and the stimulus in two peripheral apertures (each 2° wide, 28° high). To ensure that poor vergence control did not lead to a spatial overlap of stimulus and mask for dichoptic presentations,

Figure 9. Suppression from overlapping monocular and dichoptic masks. With stimulus (0.25 cpd SF, 32% contrast) and mask (0.5 cpd SF, 32% contrast) spatially overlapping, the mask is significantly more effective at suppressing the OFR induced by the stimulus (i.e., lower normalized OFR) when it is presented to the same eye (orange) than when it is presented to the other eye (blue). Black bars indicate SEM. Unpaired bootstrap-based (i.e., nonparametric) test was used to compute significance levels.

Figure 10. Suppression from nonoverlapping monocular and dichoptic masks. When stimulus (0.25 cpd SF) and mask (0.5 cpd SF, 80% contrast) are presented in a center-surround configuration (4° gap), the mask is equally effective whether is presented to the same (orange) or to the other eye (blue). In both cases, a divisive and a scaling suppressive effect of the mask are observed. Fit parameters are listed in Table 3.

we separated them by 4°. Because this is larger than Panum’s fusional area, this guarantees that, as long as the subjects perceive the fixation cross as fused, stimulus and mask will be projected on distinct retinotopic areas in the two retinæ. Because suppressive effects are much weaker in the surround configuration (Fig. 8), we only used a high-contrast mask (80%). We found (Fig. 10) that a good fit to the data required using a model that included both divisive and scaling effects, just as in Figure 8. Because only one mask contrast was used, we did not use the full models described above, but simply compared Naka–Rushton functions in which either only c_{50} or both c_{50} and A could change across the two masked (same or other eye) and the unmasked conditions (fit parameters in Table 3). Importantly, the mask was, at all contrasts of the stimulus, equally effective whether it was presented to the same or to the other eye. Because rivalry is minimal with spatially separated stimuli (Blake et al., 1992) and is thus unlikely to act as a confounding factor, these results suggest that a cortical mechanism is mostly responsible for masking suppression originating from the surround.

To summarize, it appears that, when stimulus and mask have similar SFs, masks that overlap with the stimulus (or are in its

Table 3. Fit parameters for dichoptic tuning of suppression

Figure	Subject	SF _d (cpd)	SF _m (cpd)	n	c_{50U}	c_{50MS}	c_{50MO}	A_U	A_{MS}	A_{MO}	χ^2	ΔAIC_c
10	S1	0.25	0.5	1.46	20.03	42.47	50.42	1.17	N/A	N/A	58.78	
				1.85	18.41	24.90	25.40	1.15	0.85	0.79	11.43	−34.02
	S2			1.88	12.93	30.17	32.60	0.41	N/A	N/A	32.37	
				2.68	14.47	20.98	21.17	0.45	0.31	0.30	5.71	−13.33

Alternating lines show parameters values for fits of the same dataset with Naka–Rushton functions that incorporate either a purely divisive or a divisive plus scaling suppression mechanism. In the first model, the semisaturation parameter can assume one of three values: c_{50U} in the unmasked condition, c_{50MS} when mask and stimulus are presented to the same eye, and c_{50MO} when mask and stimulus are presented one to each eye. In the second model, the gain parameter can also assume one of three values, similarly labeled. ΔAIC_c indicates the decrease in the AICc measure obtained with the divisive plus scaling model. If this value is smaller than −10, there is strong evidence to prefer this model. Otherwise, the purely divisive model should be considered sufficient to account for the data. Values that do not vary within a set of conditions are only shown on the first row.

N/A, Not applicable.

immediate surround) engage a monocular mechanism of suppression and the overall pattern of suppression appears divisive. Based on our dichoptic presentations, additional suppression at a binocular stage for overlapping stimuli can be neither proven nor excluded. In contrast, dichoptic masks and masks that are in the surround induce suppression predominantly (and possibly even exclusively) at a cortical level and their effect appears to include both divisive and scaling components.

Flickering masks: a separate mechanism?

Recently, we (Sheliga et al., 2016) reported that a flickering 1D noise stimulus superimposed on a drifting 1D pattern having the same orientation can suppress the OFRs induced by the latter; we noted that broadband noise stimuli presented on digital displays, which are affected by temporal aliasing, would engage such suppression mechanism. Here, we have shown that static masks covering a broad range of SFs can also strongly suppress OFRs. It is thus natural to ask whether there is something special about the suppression induced by flicker, as we suggested in our previous study, or if it can be seen as a special case of the suppression phenomenon that we have described here.

To explore this question, we masked a drifting sinusoidal grating (0.25 cpd, 15 Hz, 8% contrast) with a 1D noise pattern (see Materials and Methods, “Temporal filter analysis” section). The mask always had motion energy balanced in both directions (and thus induced no OFR on its own), but it could either be scattered across the entire TF spectrum (flickering mask) or concentrated within a (relatively) narrow TF bandwidth, with a center TF ranging between 0 and 45 Hz (focused mask). Importantly, focused masks are flicker free. We found (Fig. 11) that focused masks with center TF below 15 Hz provided the strongest suppression. We then identified the monophasic (i.e., low-pass) and biphasic (i.e., band-pass) temporal filters that best reproduce the strength of the suppression induced by the focused masks (Fig. 11, blue and orange lines). Finally, we used these two models to predict the strength of suppression of the flickering mask. The two filters generated almost identical predictions, which matched the measured flicker-induced suppression (right side of plots in Fig. 11). This indicates that, at least under the conditions tested, the suppression strength of a flickering stimulus is captured by the masking mechanism discussed here without having to postulate the presence of an additional nonlinearity or flicker detector.

Cascade model of contrast normalization and masking

We described above two types of suppressive effects: one that is well captured by a purely divisive model of suppression based on the Naka–Rushton function and the other that requires a Naka–

Figure 11. Suppression from nonstatic masks. Stimulus (drifting sinusoidal grating, SF = 0.25 cpd, TF = 15 Hz, contrast = 8%) and mask (high-pass-filtered 1D noise pattern) have both motion energy, but for the mask it is balanced in both directions. The TF around which the motion energy of the mask is concentrated is indicated on the abscissa. Low-TF masks are more effective at suppressing the OFR to the stimulus. Both a low-pass temporal filter (blue) and a band-pass temporal filter (orange) fit the data reasonably well. Importantly, both filters predict the amount of suppression exerted by a flickering mask (FL on the abscissa).

Rushton function that incorporates both divisive and scaling components. The former were observed only when the stimulus and the mask overlapped and had similar SFs. The latter were observed when stimulus and mask overlapped but their SFs differed considerably, when stimulus and mask did not overlap even if they had similar SF, and when stimulus and mask were presented to different eyes. These effects closely mirror those observed in cat area 17 (Sengpiel et al., 1998). Dichoptic presentations revealed that some effects are mediated mostly by monocular mechanisms, but others, especially those that result in scaling, involve binocular and thus cortical mechanisms. We now describe a simple model that is capable of capturing these various effects.

The model is structured as a cascade of two stages, loosely associated with monocular (and thus subcortical plus possibly the input layer of area V1) and binocular (and thus cortical) stages of neural processing. Each stage is modeled as a nonlinear input–output relationship described by a Naka–Rushton function with three parameters (gain, exponent, and semisaturation constant). Because the gain parameters in the two stages are redundant, in the first stage, we simply imposed a unity gain. The basic model thus has five parameters. To account for the impact of the mask, in each stage, the semisaturation constant is a function of the contrast of the mask. Indicating with y the output of the first stage and with z the output of the second, the cascade model is thus described by the following pair of equations:

$$\begin{aligned} y(c_d, c_m | SF_d, SF_m) &= \frac{c_d^{n_1}}{c_d^{n_1} + c_{50}(c_m)^{n_1}} \\ z(c_d, c_m | SF_d, SF_m) &= A_2 \frac{y^{n_2}}{y^{n_2} + y_{50}(c_m)^{n_2}} \end{aligned} \quad (7)$$

Initially, we parametrized $c_{50}(c_m)$ and $y_{50}(c_m)$ with the same quadratic functions used for the descriptive functions used above (Eq. 4), but inspection of the resulting fits led us to adopt simpler formulations. We found that the effect of the mask was well captured under all conditions tested using the following formulations:

Figure 12. Cascading effects of masking suppression. The behavior of the cascade model as a function of the relative contributions to masking suppression of the first and second stage is illustrated. Four different model instantiations are considered, differing only in the stage(s) at which the mask exerts divisive suppression: at neither stage (black), only at the first stage (gray), only at the second stage (dashed yellow), or at both stages (dashed pink). **A**, Output of the first stage. **B**, Output of the second stage.

$$c_{50}(c_m) = c_{50} + \alpha c_m$$

$$y_{50}(c_m) = y_{50} + \beta \frac{c_m}{c_m + \delta} \quad (8)$$

The overall model thus has eight parameters: n_1 and c_{50} determine the strength of contrast normalization at the first stage, α determines the strength of masking at the first stage, n_2 and y_{50} determine the strength of contrast normalization at the second stage, β and δ control masking at the second stage, and A_2 scales the overall responses. Because there are more parameters than in the descriptive functions previously fit to the data (which had five or seven parameters), this model is less well constrained and fitting it to the data requires some additional care.

Similar models have been used in the past, but because of the presence of two nonlinear contrast normalization mechanisms and two nonlinear masking mechanisms, it is far from intuitive how each parameter affects the overall behavior of the model. While exploring its properties over the entire range of parameter values is beyond the scope of this study, highlighting the behavior for a few choices of parameters can be quite useful. Accordingly, in Figure 12 we show the output of the first (Fig. 12A) and second (Fig. 12B) stage for four sets of parameter values. In all cases, we set $n_1 = 1$, $c_{50} = 10$, $A_2 = 1$, $n_2 = 3.5$, $y_{50} = 0.32$, and $\delta = 38.15$ (the reason for these choices will become clear later on). We then assigned to α and β values [0, 0] (black), [1, 0] (gray), [0, 1] (dashed yellow), and [1, 1] (dashed pink), and varied c_d between 0 and 40%, with $c_m = 10\%$. In the first instantiation of the model (black), the mask has no effect on either stage and thus only contrast normalization effects are seen, with the second stage adding to the first and thus making the relationship more saturated. These curves also show the responses of each stage for all model instantiations to the unmasked stimulus. When divisive masking is present only at the first stage (gray), a rightward shift of the curve is seen at the first stage. Because masking does not act at the second stage, a simple lateral shift is also apparent at the output of the second stage. When masking normalization occurs only at the second stage (yellow), the output of the first stage is unchanged. At the second stage, the effect is not simply to shift the relationship to the right, as expected from a divisive normalization effect; rather, a scaling effect is also present. Finally, when there is divisive masking at both stages (pink), the curve is shifted

Table 4. Fit parameters for two-stage model of suppression

Figure	Subject	SF _d (cpd)	SF _m (cpd)	n_1	c_{50}	α	A_2	n_2	y_{50}	β	δ	χ^2
3	S1	0.25	0.5	1.00	10.00	0.84	1.55	3.68	0.30	0.89	36.38	44.36
	S2					1.61	1.21	2.82	0.33	1.05	99.79	27.66
	S3					1.18	0.83	4.40	0.28	0.32	14.22	11.06
4	S1	0.125	1.0	0.75	10.00	0.00	1.54	4.15	0.51	0.88	30.27	28.19
	S3					0.01	0.86	3.94	0.43	0.85	24.97	11.04
8	S1	0.25	0.5	1.00	10.00	0.00	2.09	3.23	0.42	0.75	57.87	56.28
13	S1	0.25	0.0625	1.00	10.00	0.06	1.59	3.50	0.32	0.33	38.15	75.74
			0.125			0.33				0.29		99.00
			0.025			0.86				0.38		72.16
			0.5			0.98				0.80		45.06
			1.0			0.00				1.26		51.90
			2.0			0.00				0.38		39.59

Parameter values are for fits to Equations 7 and 8. Values that do not vary within a set of conditions are only shown on the first row.

laterally at the first stage and even further at the second. A scaling effect is also present at the output of the second stage, but it is quite small, so it might be difficult to detect with noisy measurements. This simple and physiologically realistic cascade model can thus reproduce both divisive and scaling effects despite using only divisive mechanisms. The model can thus easily account for the variety of divisive and scaling effects observed in area V1 in monkeys (Cavanaugh et al., 2002b) and cats (Sengpiel et al., 1998): cells in which strong masking effects are inherited from earlier processing would show divisive masking, whereas cells in which masking acts mostly at the cortical level would exhibit scaling effects. This of course would hold under the assumption that masking at the early levels can be modeled as a divisive interaction and that the contrast input–output function of the first stage is compressive. As noted in the introduction, abundant experimental evidence supports these assumptions within the magnocellular pathway of primates. However, describing the effect of the mask as divisive should not be taken to imply that, at the neuronal level, a divisive mechanism (e.g., shunting inhibition) is at play because many alternatives are possible (Murphy and Miller, 2003; Priebe and Ferster, 2012).

Armed with an understanding of the types of behaviors that can be expected from the model, we can now use it to fit the data from our experiments. We started by fitting the effects of masking observed with overlapping stimulus and mask, shown in Figures 3 and 4. Because the model is not strongly constrained, we first computed best fits for a wide range of imposed values of n_1 and c_{50} . We found that similarly good fits could be obtained over a range of these parameters. By intersecting these regions across subjects and experiments, we found that imposing $c_{50} = 10$ worked well in all cases; furthermore, using $n_1 = 1$ worked well when $SF_d = 0.25$ cpd, whereas $n_1 = 0.75$ was appropriate when $SF_d = 0.125$ cpd. With these two parameters so selected, the model fits were well constrained and excellent in all cases, usually indistinguishable from those obtained with the descriptive functions (so we chose not to show them in a separate figure). Parameter values and quality of fit measures are reported in Table 4.

To fully explore the range of effects that can be accounted for by the model, in one subject, we measured the effect of the overlapping mask across the full range of stimulus and mask contrasts for one SF of the stimulus (0.25 cpd) and six SFs of the mask (0.0625, 0.125, 0.25, 0.5, 1.0, and 2.0 cpd). Because, within each session, only a single mask SF was used, we collected six times as much data for each unmasked condition as for each masked condition. In fitting the model to the data, we thus first imposed $n_1 =$

1 and $c_{50} = 10$ and fitted A_2 , n_2 , and y_{50} to the unmasked data; with these parameters so selected, we then found, for each SF of the mask, the best fitting values for α , β , and δ . To further reduce the degrees of freedom, we then selected the value of δ from the dataset in which the mask exerts the strongest influence on the second stage ($c_m = 1.0$ cpd) and fit all datasets once again with this value for δ (which resulted in only marginally inferior fits). Data and model fits are shown in Figure 13A, with parameter values and quality of fit measures reported in Table 4. In Figure 13B, we show how the only two parameters that differed across datasets (α and β) vary as a function of the SF of the mask, together with log-Gaussian fits to this relationship ($R^2 = 0.934$ for first stage, $R^2 = 0.994$ for second stage). Note that the second stage is subject to stronger masking effects at higher SFs ($\beta_{\text{peak}} = 0.81$ cpd) relative to the first stage ($\alpha_{\text{peak}} = 0.34$ cpd). At the lowest and highest SF, the mask had no effect on the first stage, whereas some masking was evident at all SFs in the second stage ($\beta_{\text{min}} = 0.33$). The SF spectrum over which the parameters were modulated significantly was also different for the two stages, being broader for the first stage ($\alpha_{\text{BW}} = 1.86$ and $\beta_{\text{BW}} = 1.25$, both expressed in octaves). Because this result might be contingent on the particular values selected for n_1 , c_{50} , and δ , we repeated the entire optimization process sampling a wide range of values for n_1 , c_{50} . We found that, whenever the final overall fits were acceptable [i.e., the mean AICc across the six mask SFs did not increase by >10 points, a lenient criterion, meaning that significantly worse fits than those obtained with our initial choice for (n_1 , c_{50}) were included], α and β were always well fit by log-Gaussian functions ($R^2 = 0.93 \pm 0.03$ for the first and $R^2 = 0.99 \pm 0.01$ for the second stage), closely resembling those shown in Figure 13B ($\alpha_{\text{min}} = 0.04 \pm 0.06$, $\alpha_{\text{peak}} = 0.36 \pm 0.02$, $\alpha_{\text{BW}} = 1.77 \pm 0.25$, $\beta_{\text{min}} = 0.26 \pm 0.08$, $\beta_{\text{peak}} = 0.84 \pm 0.04$, $\beta_{\text{BW}} = 1.35 \pm 0.07$). This result is thus robust relative to the particular choice of parameters and is a feature of the model (and possibly of the data to which it is applied). It must be stressed that the function $\beta(\text{SF})$ shown in Figure 13B illustrates the SF tuning of masking of the second stage in isolation without taking into account the effects that the mask has on the output of the first stage, so it is quite different from the SF tuning of masking of the visual input measured at the output of the second stage, which is shown in Figure 2. Importantly, it is the latter that would be measured if one were to “record” from the output of the second stage, as is done in neurophysiological experiments.

Figure 12 provided a general idea about the conditions under which scaling effects emerge. To better understand when scaling is significant enough to be detected reliably, we simulated the behavior of the cascade model over a wide range of values for α and β , with all the other parameters as in Figure 13 (which were also used in Fig. 12). From each simulation, we generated synthetic data at the same signal and mask contrast levels used in the experiment, with the average noise level observed in the experiments and fit to it the divisive and divisive plus scaling functions used earlier (Eqs. 3 and 5). In Figure 14, we plot the difference between the χ^2 measures for the two functions, a measure of the extent by which the fit was improved by the addition of the scaling component. The region over which the improvement is significant based on the AICc measure is outlined in orange. Blue lower-case letters indicate the parameter values associated with each of the six subpanels in Figure 13A. When SF of stimulus and mask are separated by two or more octaves, a significant scaling component of suppression can be detected; for smaller SF separations, the presence of such an effect cannot be demonstrated statistically and a purely divisive model suffices. As a first-order

Figure 13. Modeling contrast tuning of masking suppression over a wide range of mask SFs. **A**, Same as Figures 3 and 4, but here for a single SF of the stimulus (0.25 cpd) the contrast tuning of suppression is measured for a wide range of mask SFs (indicated in each subpanel). The data are fit using a two-stage cascade model. The model has eight parameters, but only two are different in each of the six panels, one in the first stage (α) and one in the second stage (β). Therefore, all of the fits shown here require a combined total of 18 parameters. Parameters for the fits are listed in Table 4. **B**, α (top) and β (bottom) vary in a lawful manner as a function of the SF of the mask. Log-Gaussian fits are shown.

Figure 14. Divisive and scaling components of suppression. The values of parameters α (abscissa) and β (ordinate) in the cascade model (Eq. 8) predict whether the effect of suppression can be described by a purely divisive model or if a scaling component of suppression must also be included. Here, a wide range of values for both parameters is explored using simulated data and the increase in the χ^2 measure associated with the purely divisive model is shown as a grayscale heatmap. The region over which the improvement is significant according to the AICc is outlined (orange). Lowercase letters are placed in correspondence to the parameter values for the fits in the subpanels in Figure 13A. *a*, 0.0625 cpd; *b*, 0.125 cpd; *c*, 0.25 cpd; *d*, 0.5 cpd; *e*, 1.0 cpd; *f*, 2.0 cpd. Mask SFs two or more octaves away from the stimulus SF required a scaling component, whereas closer ones did not.

approximation, a significant scaling effect is detected whenever β is larger than ~ 0.3 and α is smaller than 0.5.

We can also use the model to fit the data from nonoverlapping stimulus and mask, shown in Figure 8. Once again, the fits were excellent and slightly better than those obtained with the descriptive functions. Parameter values and quality of fit measures are reported in Table 4. Note that, whereas for overlapping stimuli (with the same SFs for stimulus and mask), α was quite large, now it is zero, indicating that the mask does not act at the first stage of the model when mask and stimulus do not overlap (at least for the combination of stimulus and mask SF tested here).

Unfortunately, we do not have enough data to characterize fully the tuning properties of the second stage with surround masking (i.e., how β varies with mask SF). Nevertheless, if the mask were to act only at the second stage for all SFs, then the SF tuning of β at the second stage would simply mirror the SF tuning of the overall masking effect, as shown in Figure 7. It would thus be quite different from the SF tuning for overlapping masks (Fig. 13B): it would be tuned to a lower SF, much closer to the SF of the stimulus, and would have a much broader bandwidth (tuning parameters can be found in Table 1).

Another potentially interesting feature of the model relates to the signals used to normalize the response divisively at the two stages (Eq. 8). At the first stage, this signal is directly proportional to the mask contrast; at the second stage, it is proportional to a compressive function of the mask contrast. This tentatively suggests that, in subcortical magnocellular neurons, suppression might originate from photoreceptors and/or from parvocellular

signals (the responses of which are linearly related to stimulus contrast); in cortical binocular neurons, magnocellular signals (which nonlinearly encode contrast) might instead contribute to the masking effect.

Our simple model is not meant to simulate the behavior of individual neurons in visual cortex and thus it embeds no RF concept. However, it would be straightforward to extend it in such a way, with stimulus contrast and mask contrast at the two stages being measured over defined regions of retinal space (i.e., RFs) and the second stage pooling across a subpopulation of first-stage outputs. The differences that we reported above between overlapping and center-surround masks would then readily emerge in a model in which at the first-stage mask and stimulus are measured over the same area, whereas at the second stage, the RF for the mask is larger than that for the stimulus.

Of course, other models could also account for our data. For example, mask and stimulus could be processed separately and their contrast and speed (zero for the mask) could be used to decide whether to move the eyes (i.e., follow the stimulus) or to keep fixating (i.e., follow the mask). This scheme would be especially applicable to the condition in which stimulus and mask are spatially separate and, as our binocular experiments revealed, the stimulus–mask interactions are mostly cortically mediated. This alternative interpretation, although conceptually different, might rely on an implementation that is similar to the simpler account that we propose and does not make distinct predictions for the experiments presented here.

Discussion

We characterized the effect of a static mask on the OFRs induced by a moving stimulus at the short time scale typical of intersaccadic fixation periods. We found that, even when mask and stimulus have similar contrasts, the mask can almost obliterate the OFR. When mask and stimulus overlap in space and have similar SFs, the effect of the mask is purely divisive. However, when they differ considerably in SF or when stimulus and mask do not

overlap or are presented dichoptically, a scaling component of suppression appears.

To gain insights into the underlying processes, we fit a two-stage model to our data. Each stage of the model incorporates a nonlinear input–output function and divisive contrast masking. The model revealed that suppression appears divisive whenever the mask strongly affects the first stage. Furthermore if the first stage has a compressive input–output contrast response function but is weakly affected by the static mask, masking at the second stage produces scaling effects. Based on dichoptic masking experiments, we mapped the first stage to areas processing monocular signals (retina, LGN, and possibly the first layer of area V1) and the second stage to binocular areas (areas V1, MT, and MST). We then inferred that the divisive effects induced by masks that are overlapping and of similar scale are mostly determined by divisive interactions occurring at subcortical levels, although the mask might further act at cortical levels. Furthermore, scaling effects observed with nonoverlapping masks, dichoptic masks, or masks of a widely different SF seem to be mediated mostly by cortical divisive phenomena more likely to occur within area V1 than in later areas (based on RF size considerations).

Neurophysiological investigations of suppression of cortical and subcortical neuronal responses by masks placed either within the CRF of neurons or in the surrounding region have a long history. Studies in the primary visual cortex of cats and monkeys revealed that suppression from the CRF is mostly monocular (DeAngelis et al., 1992), not tuned for orientation (Bonds, 1989; DeAngelis et al., 1992; Carandini et al., 1997), and likely contributes to contrast gain control. At least in the cat, it is broadly tuned for TF (Allison et al., 2001; Freeman et al., 2002) and SF, with a tendency to be stronger when mask SF is lower than the preferred SF of the excitatory mechanism (Morrone et al., 1982; Bonds, 1989; DeAngelis et al., 1992). Overlay suppression is thought to arise mostly in subcortical structures such as the retinal ganglion cells and the LGN (Freeman et al., 2002; Li et al., 2006; Priebe and Ferster, 2006), with cortical neurons contributing to it mostly by virtue of their output nonlinearities (Albrecht and Geisler, 1991). Therefore, studies in cortex might have mostly measured masking effects inherited from LGN. In both cats and monkeys, surround suppression measured in primary visual cortex behaves quite differently: is tuned for both orientation and SF, being strongest when the center and surround have the same orientation and SF (DeAngelis et al., 1994), even when they differ from the preferred values for the cell (Cavanaugh et al., 2002b), and is similarly strong monocularly and dichoptically (DeAngelis et al., 1994). It originates mostly in cortex, possibly through feedback from extrastriate areas to V1 (Cavanaugh et al., 2002a), although it might have a subcortical component as well (Solomon et al., 2006) mediated mostly by feedback from V1 to LGN (Webb et al., 2002; Nolt et al., 2007). Both overlay and surround suppression arise very quickly either together with or within 10–20 ms from the onset of the excitatory response (Albrecht et al., 2002; Bair et al., 2003; Smith et al., 2006). Overlay suppression can be modeled as a divisive phenomenon (Carandini et al., 1997), whereas surround suppression combines divisive and scaling effects (Sengpiel et al., 1998; Cavanaugh et al., 2002b).

Our results and the inferences that we drew from the model are largely compatible with these data, but there are some apparent discrepancies. With an overlapping mask, our model infers a strong contribution of the mask at the first stage, in agreement with the general consensus about the strong contribution of subcortical processing to overlay masking. Masking at the first stage in our model is also fairly tightly tuned around the SF of the

stimulus (Fig. 13B). This is at odds with the broad SF tuning of LGN neurons (which presumably provide the signal for monocular overlay masking) to small stimuli; however, when studied with large stimuli, most LGN neurons have a band-pass SF tuning (Derrington and Lennie, 1984; Solomon et al., 2002; Bonin et al., 2005; Alitto and Usrey, 2008). The SF tuning of overlay suppression has also been reported to be fairly broad in the primary visual area (DeAngelis et al., 1992) and when tested psychophysically (Petrov et al., 2005). However, in these studies, the contrast of the mask was considerably higher than the contrast of the stimulus and the width of SF tuning for masking increases with the ratio between the two, as demonstrated here (Fig. 5) and in cat area 17 (Bonds, 1989). Accordingly, the response of magnocellular LGN neurons to large transient stimuli might be little affected by a low-frequency static mask of equal contrast.

With nonoverlapping masks, our model infers that masking mostly occurs at the second stage, which is again consistent with neurophysiological evidence. Once again, the SF tuning properties do not match perfectly the neurophysiological (Cavanaugh et al., 2002b) or psychophysical (Petrov et al., 2005) data, which find surround suppression to be tightly tuned around the SF of the driving stimulus. In our data, for nonoverlapping stimuli, the most effective mask SF is independent of the SF of the drifting stimulus (Fig. 7), although the range that we tested is quite limited and the deviation was only half an octave. The bandwidth was instead consistent with the existing data. Any of the many differences between the spatiotemporal properties of the stimuli used in this study and those normally used in neurophysiological and psychophysical studies could be responsible for these small differences.

OFRs are not the only eye movements affected by the presence of a static mask. Although steady-state tracking of a small target is marginally affected by the presence of a static textured background, the initial acceleration can be strongly attenuated by it (Collewijn and Tamminga, 1984; Keller and Khan, 1986; Kimmig et al., 1992; Masson et al., 1995; Mohrmann and Thier, 1995; Niemann and Hoffmann, 1997). The attenuation is reduced by 50% when target and background are presented dichoptically, indicating that at least part of it is of cortical origin. These similarities, along with the observations reported here, are perhaps not surprising because smooth pursuit and ocular following may share the same neural substrate (dorsal visual pathway). However, there are also some important differences. In particular, we found that, as soon as stimulus and mask are spatially offset, the suppression drops markedly; this is not the case with smooth pursuit, in which separating target and background by 2° leads to only a small decrease in attenuation (Kimmig et al., 1992). Significant differences also exist when comparing our results with the effect of static masks on motion perception. Previous studies reported that, with short exposures, the presence of an overlapping static mask can reverse the perceived direction of motion of the stimulus, especially when mask SF is lower than stimulus SF (Derrington and Henning, 1987, 1989; Henning and Derrington, 1988; Serrano-Pedraza et al., 2007). We did not observe OFR reversals and found stronger effects for overlapping masks of somewhat higher SFs than the stimulus (Fig. 2B). Discrepancies between eye movements and perception are however not novel (Boström and Warzecha, 2010; Simoncini et al., 2012; Blum and Price, 2014; Glasser and Tadin, 2014; Price and Blum, 2014; Quaia et al., 2016) and might reflect a different weighting of visual signals carried by various areas/pathways.

Although, under some conditions, we have the ability to perceive and process multiple stimuli simultaneously and inde-

pendently, many of our most important actions involve binary choices: should we freeze or run in response to a threat? The stabilization of our gaze, crucial to our ability to process visual information, falls under this second category: Should we stabilize on the retina a static target or a moving one? Static, flickering, and slow motion signals thus have to compete with fast motion signals and we showed that they can reduce dramatically the eye response that the latter would elicit in isolation. Our results reveal that this process occurs within a well defined spatiotemporal spectral region and is strongly determined by the relative contrast of the various stimuli present. The neuronal circuitry required to perform this balancing act might be surprisingly simple and involve suppression mechanisms that have long been known to operate in subcortical and cortical visual areas.

References

- Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2:284–299. [CrossRef Medline](#)
- Albrecht DG, Geisler WS (1991) Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Vis Neurosci* 7:531–546. [CrossRef Medline](#)
- Albrecht DG, Geisler WS, Frazor RA, Crane AM (2002) Visual cortex neurons of monkeys and cats: temporal dynamics of the contrast response function. *J Neurophysiol* 88:888–913. [Medline](#)
- Alitto HJ, Usrey WM (2008) Origin and dynamics of extraclassical suppression in the lateral geniculate nucleus of the macaque monkey. *Neuron* 57:135–146. [CrossRef Medline](#)
- Alitto HJ, Moore BD 4th, Rathbun DL, Usrey WM (2011) A comparison of visual responses in the lateral geniculate nucleus of alert and anesthetized macaque monkeys. *J Physiol* 589:87–99. [CrossRef Medline](#)
- Allison JD, Smith KR, Bonds AB (2001) Temporal-frequency tuning of cross-orientation suppression in the cat striate cortex. *Vis Neurosci* 18:941–948. [Medline](#)
- Bair W, Cavanaugh JR, Movshon JA (2003) Time course and time-distance relationships for surround suppression in macaque V1 neurons. *J Neurosci* 23:7690–7701. [Medline](#)
- Barthélemy FV, Vanzetta I, Masson GS (2006) Behavioral receptive field for ocular following in humans: dynamics of spatial summation and center-surround interactions. *J Neurophysiol* 95:3712–3726. [CrossRef Medline](#)
- Bedwell JS, Brown JM, Miller LS (2003) The magnocellular visual system and schizophrenia: what can the color red tell us? *Schizophr Res* 63:273–284. [CrossRef Medline](#)
- Bedwell JS, Miller LS, Brown JM, Yanasak NE (2006) Schizophrenia and red light: fMRI evidence for a novel biobehavioral marker. *Int J Neurosci* 116:881–894. [CrossRef Medline](#)
- Benardete EA, Kaplan E, Knight BW (1992) Contrast gain control in the primate retina: P cells are not X-like, some M cells are. *Vis Neurosci* 8:483–486. [CrossRef Medline](#)
- Blake R, O'Shea RP, Mueller TJ (1992) Spatial zones of binocular rivalry in central and peripheral vision. *Vis Neurosci* 8:469–478. [CrossRef Medline](#)
- Blum J, Price NS (2014) Reflexive tracking eye movements and motion perception: one or two neural populations? *J Vis* 14:23. [CrossRef Medline](#)
- Bonds AB (1989) Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Vis Neurosci* 2:41–55. [CrossRef Medline](#)
- Bonin V, Mante V, Carandini M (2005) The suppressive field of neurons in lateral geniculate nucleus. *J Neurosci* 25:10844–10856. [CrossRef Medline](#)
- Boström KJ, Warzecha AK (2010) Open-loop speed discrimination performance of ocular following response and perception. *Vision Res* 50:870–882. [CrossRef Medline](#)
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436. [CrossRef Medline](#)
- Breitmeyer BG, Breier JI (1994) Effects of background color on reaction time to stimuli varying in size and contrast: inferences about human m channels. *Vision Res* 34:1039–1045. [CrossRef Medline](#)
- Breitmeyer BG, Williams MC (1990) Effects of isoluminant-background color on metacontrast and stroboscopic motion: interactions between sustained (p) and transient (m) channels. *Vision Res* 30:1069–1075. [CrossRef Medline](#)
- Burnham KP, Anderson RA (2002) Model selection and multimodal inference. New York: Springer.
- Camp AJ, Tailby C, Solomon SG (2009) Adaptable mechanisms that regulate the contrast response of neurons in the primate lateral geniculate nucleus. *J Neurosci* 29:5009–5021. [CrossRef Medline](#)
- Carandini M, Heeger DJ, Movshon JA (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. *J Neurosci* 17:8621–8644. [Medline](#)
- Cavanaugh JR, Bair W, Movshon JA (2002a) Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *J Neurophysiol* 88:2530–2546. [CrossRef Medline](#)
- Cavanaugh JR, Bair W, Movshon JA (2002b) Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. *J Neurophysiol* 88:2547–2556. [CrossRef Medline](#)
- Chapman C, Hoag R, Giaschi D (2004) The effect of disrupting the human magnocellular pathway on global motion perception. *Vision Res* 44:2551–2557. [CrossRef Medline](#)
- Collewijn H, Tamminga EP (1984) Human smooth and saccadic eye movements during voluntary pursuit of different target motions on different backgrounds. *J Physiol* 351:217–250. [CrossRef Medline](#)
- Collewijn H, van der Mark F, Jansen TC (1975) Precise recording of human eye movements. *Vision Res* 15:447–450. [CrossRef Medline](#)
- David SV, Gallant JL (2005) Predicting neuronal responses during natural vision. *Network* 16:239–260. [CrossRef Medline](#)
- DeAngelis GC, Newsome WT (2004) Perceptual 'read-out' of conjoined direction and disparity maps in extrastriate area MT. *PLoS Biol* 2:E77. [CrossRef Medline](#)
- DeAngelis GC, Robson JG, Ohzawa I, Freeman RD (1992) Organization of suppression in receptive fields of neurons in cat visual cortex. *J Neurophysiol* 68:144–163. [Medline](#)
- DeAngelis GC, Freeman RD, Ohzawa I (1994) Length and width tuning of neurons in the cat's primary visual cortex. *J Neurophysiol* 71:347–374. [Medline](#)
- Derrington AM, Henning GB (1987) Errors in direction-of-motion discrimination with complex stimuli. *Vision Res* 27:61–75. [CrossRef Medline](#)
- Derrington AM, Henning GB (1989) Some observations on the masking effects of two-dimensional stimuli. *Vision Res* 29:241–246. [CrossRef Medline](#)
- Derrington AM, Lennie P (1984) Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *J Physiol* 357:219–240. [CrossRef Medline](#)
- Dreher B, Fukada Y, Rodieck RW (1976) Identification, classification and anatomical segregation of cells with x-like and y-like properties in the lateral geniculate nucleus of old-world primates. *J Physiol* 258:433–452. [CrossRef Medline](#)
- Edwards VT, Hogben JH, Clark CD, Pratt C (1996) Effects of a red background on magnocellular functioning in average and specifically disabled readers. *Vision Res* 36:1037–1045. [CrossRef Medline](#)
- Efron, B (1982) The jackknife, the bootstrap, and other resampling plans. Philadelphia: SIAM.
- Freeman TC, Durand S, Kiper DC, Carandini M (2002) Suppression without inhibition in visual cortex. *Neuron* 35:759–771. [CrossRef Medline](#)
- Geisler WS, Albrecht DG, Crane AM (2007) Responses of neurons in primary visual cortex to transient changes in local contrast and luminance. *J Neurosci* 27:5063–5067. [CrossRef Medline](#)
- Glasser DM, Tadin D (2014) Modularity in the motion system: independent oculomotor and perceptual processing of brief moving stimuli. *J Vis* 14:1–13. [CrossRef](#)
- Hays, AV, BJ Richmond, LM Optican (1982) A UNIX-based multiple process system for real-time data acquisition and control. Presented at the WESCON Conference, El Segundo, CA, September 14–16.
- Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9:181–197. [CrossRef Medline](#)
- Henning GB, Derrington AM (1988) Direction-of-motion discrimination with complex patterns: further observations. *J Opt Soc Am A* 5:1759–1766. [CrossRef Medline](#)
- Kaplan E, Shapley RM (1986) The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. *Proc Natl Acad Sci U S A* 83:2755–2757. [CrossRef Medline](#)
- Kawano K (1999) Ocular tracking: behavior and neurophysiology. *Curr Opin Neurobiol* 9:467–473. [CrossRef Medline](#)

- Keller EL, Khan NS (1986) Smooth-pursuit initiation in the presence of a textured background in monkey. *Vision Res* 26:943–955. [CrossRef Medline](#)
- Kimmig HG, Miles FA, Schwarz U (1992) Effects of stationary textured backgrounds on the initiation of pursuit eye movements in monkeys. *J Neurophysiol* 68:2147–2164. [Medline](#)
- Kremers J, Silveira LC, Kilavik BE (2001) Influence of contrast on the responses of marmoset lateral geniculate cells to drifting gratings. *J Neurophysiol* 85:235–246. [Medline](#)
- Krüger J (1977) Stimulus dependent colour specificity of monkey lateral geniculate neurones. *Exp Brain Res* 30:297–311. [Medline](#)
- Li B, Thompson JK, Duong T, Peterson MR, Freeman RD (2006) Origins of cross-orientation suppression in the visual cortex. *J Neurophysiol* 96:1755–1764. [CrossRef Medline](#)
- Masson GS (2004) From 1D to 2D via 3D: dynamics of surface motion segmentation for ocular tracking in primates. *J Physiol Paris* 98:35–52. [CrossRef Medline](#)
- Masson GS, Perrinet LU (2012) The behavioral receptive field underlying motion integration for primate tracking eye movements. *Neurosci Biobehav Rev* 36:1–25. [CrossRef Medline](#)
- Masson G, Proteau L, Mestre DR (1995) Effects of stationary and moving textured backgrounds on the visuo-oculo-manual tracking in humans. *Vision Res* 35:837–852. [CrossRef Medline](#)
- Miles FA (1997) Visual stabilization of the eyes in primates. *Curr Opin Neurobiol* 7:867–871. [CrossRef Medline](#)
- Miles FA (1998) The neural processing of 3-D visual information: evidence from eye movements. *Eur J Neurosci* 10:811–822. [CrossRef Medline](#)
- Miles, FA and BM Sheliga (2010) Motion detection for reflexive tracking. In: *Dynamics of visual motion processing: neuronal, behavioral, and computational approaches* (Ilg, UJ and GS Masson, eds), pp 141–160. New York: Springer.
- Miles FA, Kawano K, Optican LM (1986) Short-latency ocular following responses of monkey. I. Dependence on temporospatial properties of visual input. *J Neurophysiol* 56:1321–1354. [Medline](#)
- Miles, FA, C Busetini, GS Masson, and DS Yang (2004) Short-latency eye movements: evidence for rapid, parallel processing of optic flow. In: *Optic flow and beyond* (Vaina, LM, Beardsley SA, Rushton S, eds), pp 79–107. Dordrecht, The Netherlands: Kluwer Academic.
- Miura K, Matsuura K, Taki M, Tabata H, Inaba N, Kawano K, Miles FA (2006) The visual motion detectors underlying ocular following responses in monkeys. *Vision Res* 46:869–878. [CrossRef Medline](#)
- Mohrmann H, Thier P (1995) The influence of structured visual backgrounds on smooth-pursuit initiation, steady-state pursuit and smooth-pursuit termination. *Biol Cybern* 73:83–93. [CrossRef Medline](#)
- Morrone MC, Burr DC, Maffei L (1982) Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. *Proc R Soc Lond B Biol Sci* 216:335–354. [CrossRef Medline](#)
- Murphy BK, Miller KD (2003) Multiplicative gain changes are induced by excitation or inhibition alone. *J Neurosci* 23:10040–10051. [Medline](#)
- Niemann T, Hoffmann KP (1997) The influence of stationary and moving textured backgrounds on smooth-pursuit initiation and steady state pursuit in humans. *Exp Brain Res* 115:531–540. [CrossRef Medline](#)
- Nishimoto S, Gallant JL (2011) A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *J Neurosci* 31:14551–14564. [CrossRef Medline](#)
- Nolt MJ, Kumbhani RD, Palmer LA (2007) Suppression at high spatial frequencies in the lateral geniculate nucleus of the cat. *J Neurophysiol* 98:1167–1180. [CrossRef Medline](#)
- Pammer K, Lovegrove W (2001) The influence of color on transient system activity: implications for dyslexia research. *Percept Psychophys* 63:490–500. [CrossRef Medline](#)
- Pelli DG, Zhang L (1991) Accurate control of contrast on microcomputer displays. *Vision Res* 31:1337–1350. [CrossRef Medline](#)
- Petrov Y, Carandini M, McKee S (2005) Two distinct mechanisms of suppression in human vision. *J Neurosci* 25:8704–8707. [CrossRef Medline](#)
- Price NS, Blum J (2014) Motion perception correlates with volitional but not reflexive eye movements. *Neuroscience* 277:435–445. [CrossRef Medline](#)
- Priebe NJ, Ferster D (2006) Mechanisms underlying cross-orientation suppression in cat visual cortex. *Nat Neurosci* 9:552–561. [CrossRef Medline](#)
- Priebe NJ, Ferster D (2012) Mechanisms of neuronal computation in mammalian visual cortex. *Neuron* 75:194–208. [CrossRef Medline](#)
- Quaia C, Sheliga BM, Fitzgibbon EJ, Optican LM (2012) Ocular following in humans: spatial properties. *J Vis* 12: pii: 13. [CrossRef Medline](#)
- Quaia C, Optican LM, Cumming BG (2013) Terminator disparity contributes to stereo matching for eye movements and perception. *J Neurosci* 33:18867–18879. [CrossRef Medline](#)
- Quaia C, Optican LM, Cumming BG (2016) A motion-from-form mechanism contributes to extracting pattern motion from plaids. *J Neurosci* 36:3903–3918. [CrossRef Medline](#)
- Robinson DA (1963) A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Trans Biomed Eng* 10:137–145. [Medline](#)
- Sclar G, Maunsell JH, Lennie P (1990) Coding of image contrast in central visual pathways of the macaque monkey. *Vision Res* 30:1–10. [CrossRef Medline](#)
- Sengpiel F, Baddeley RJ, Freeman TC, Harrad R, Blakemore C (1998) Different mechanisms underlie three inhibitory phenomena in cat area 17. *Vision Res* 38:2067–2080. [CrossRef Medline](#)
- Serrano-Pedraza I, Goddard P, Derrington AM (2007) Evidence for reciprocal antagonism between motion sensors tuned to coarse and fine features. *J Vis* 7:8.1–14. [Medline](#)
- Sheliga BM, Chen KJ, Fitzgibbon EJ, Miles FA (2005) Initial ocular following in humans: a response to first-order motion energy. *Vision Res* 45:3307–3321. [CrossRef Medline](#)
- Sheliga BM, Quaia C, FitzGibbon EJ, Cumming BG (2016) Ocular-following responses to white noise stimuli in humans reveal a novel non-linearity that results from temporal sampling. *J Vis* 16:8. [CrossRef Medline](#)
- Simoncini C, Perrinet LU, Montagnini A, Mamassian P, Masson GS (2012) More is not always better: adaptive gain control explains dissociation between perception and action. *Nat Neurosci* 15:1596–1603. [CrossRef Medline](#)
- Smith MA, Bair W, Movshon JA (2006) Dynamics of suppression in macaque primary visual cortex. *J Neurosci* 26:4826–4834. [CrossRef Medline](#)
- Solomon SG, White AJ, Martin PR (2002) Extraclassical receptive field properties of parvocellular, magnocellular, and koniocellular cells in the primate lateral geniculate nucleus. *J Neurosci* 22:338–349. [Medline](#)
- Solomon SG, Lee BB, Sun H (2006) Suppressive surrounds and contrast gain in magnocellular-pathway retinal ganglion cells of macaque. *J Neurosci* 26:8715–8726. [CrossRef Medline](#)
- Takemura A, Kawano K (2002) Sensory-to-motor processing of the ocular-following response. *Neurosci Res* 43:201–206. [CrossRef Medline](#)
- Takemura A, Murata Y, Kawano K, Miles FA (2007) Deficits in short-latency tracking eye movements after chemical lesions in monkey cortical areas MT and MST. *J Neurosci* 27:529–541. [CrossRef Medline](#)
- Webb BS, Tinsley CJ, Barraclough NE, Easton A, Parker A, Derrington AM (2002) Feedback from V1 and inhibition from beyond the classical receptive field modulates the responses of neurons in the primate lateral geniculate nucleus. *Vis Neurosci* 19:583–592. [CrossRef Medline](#)
- Wiesel TN, Hubel DH (1966) Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *J Neurophysiol* 29:1115–1156. [Medline](#)
- Willmore BD, Prenger RJ, Gallant JL (2010) Neural representation of natural images in visual area V2. *J Neurosci* 30:2102–2114. [CrossRef Medline](#)