INTRODUCTION

Receptive field (RF) sizes vary considerably across extrastriate cortical areas in primates. The RF size of an area often correlates well with its position on anatomically defined cortical hierarchies (Van Essen et al. 1990). This trend is well exemplified in the well-studied “motion system” that connects V1 to parietal cortical areas. An intermediate structure on this pathway, MT, has RFs of approximately 100 times greater area than those in V1 (Van Essen et al. 1981). Because the area of MT RFs is substantially larger than that of their inputs, MT cells must accumulate signals from multiple V1 cells overlapping the MT RF. The mechanisms of spatial summation have been extensively studied at earlier levels of the visual system, and these studies have been very revealing about RF mechanisms. Therefore studying the mechanisms of spatial summation in extrastriate cortex may prove similarly revealing regarding RF structure and mechanisms.

In previous work, it has been demonstrated that MT cells do not sum their inputs linearly. When presented with multiple stimuli within the RF, MT cells typically give a response much less than that expected from summing the response of the component stimuli (Britten and Heuer 1999; Ferrera and Lisberger 1997; Recanzone et al. 1997). This response re-scaling or normalization is computationally useful for two reasons. First, it keeps the response from saturating and becoming accordingly less informative. Second, response normalization removes the effects of changes in overall stimulus contrast, allowing cells to signal more meaningful aspects of the scene such as direction or speed of motion. Formal models that include such a contrast-dependent normalization step account for a wide range of physiological observations from MT (Simoncelli and Heeger 1998).

In the present experiments, we sought to investigate the mechanism of contrast normalization in MT by quantitatively characterizing its contrast dependence. To do this, we presented local stimuli within the RF of MT cells; stimuli varied in both location and contrast. These stimuli were presented in rapid succession, either singly or in pairs, and the responses to pairs of stimuli were compared against the responses to the single stimuli of which the pairs were composed. We found that normalization became effective at quite low contrasts, ones that provoked less than half-maximal responses to single stimuli. This high-contrast sensitivity of normalization (modestly higher than the contrast sensitivity of responses to single stimuli) suggests that multiple stimuli cooperate to normalize the responses of an MT cell. In our analysis, we modeled this cooperativity as a multiplicative interaction term, responsible for a divisive re-scaling of the excitatory responses. We found that this model provided good account of our data.

METHODS

Preparation

Three adult female rhesus macaques (Macaca mulatta) were used in this study. Prior to recording, each monkey was implanted with a scleral search coil (Judge et al. 1980) to monitor eye position and trained to fixate small stationary targets in the presence of visual stimuli. Additionally, each was implanted with a stainless steel head-restraint post and recording chamber located over occipital cortex. A plastic grid coordinate system placed within the recording chamber provided guide tube support at 1-mm intervals (Crist et al. 1988). For recording sessions, a stainless steel transdural guide tube was inserted at known locations within this grid. A parylene-coated tungsten microelectrode (MicroProbe) was introduced through the guide tube and
advanced using a hydraulic stepping motor (National Aperture). We used both physiological and anatomical landmarks to localize area MT. Anatomical landmarks included recording depth, gray-white matter transitions, and passages through the lumen of the superior temporal sulcus (STS). Physiologically, we required a preponderance of directionally selective neurons (Dubner and Zeki 1971), appropriate (RF) size (Maunsell and Van Essen 1983b), systematic changes in preferred direction (Albright et al. 1984), and the expected retinotopy (Van Essen et al. 1981). Our application of these criteria was conservative; if there was any doubt as to being within area MT, the data were not used for this study. Histological verification was obtained for one monkey from the study, which confirmed that the recording area was in the densely myelinated region on the posterior bank of the STS, corresponding to the normal location for area MT. The other two monkeys are still alive and being used in related experiments.

After we localized MT, we would isolate and record single-unit activity using standard extracellular techniques. Electrode signals were amplified and filtered, and single units were isolated with a window-discriminator (Bak Electronics), and their action potentials converted to TTL pulses. We used the public-domain software package REX (Hayes et al. 1982) to record the time of stimulus events and action potentials with 1-ms resolution. Once a unit was isolated, we determined the RF size, location, and preferred direction qualitatively using handheld moving bar stimuli or computer-generated moving Gabor patches and then started quantitative testing.

**Stimuli**

All stimuli were presented on the face of a CRT monitor, subtending 40° horizontally by 30° vertically at a viewing distance of 57 cm from the monkey. Pixel resolution was 1280 × 1024, with a vertical refresh rate of 72 Hz, corresponding to a frame rate of 13.9 ms. The stimuli were generated using custom software on a Pentium computer with a video card (ATI Mach 64) set to provide 8-bit grayscale resolution. Mean screen luminance was set to 30 cd/M², with a background luminance of 0.1 cd/M². The monitor was regularly calibrated, and stimuli were generated using a linearized lookup table.

The stimuli for these experiments were small moving oriented two-dimensional Gabor patches, effectively “motion impulses.” The contrast of each stimulus was a trapezoidal function over the seven frame stimulus duration (98 ms), rising on the first two frames, a constant maximal (designated) contrast for the intermediate three, and falling on the last two frames. The temporal parameters of the stimuli were constant, but spatial frequency, dimensions, and drift rate were varied across experiments. The default settings were used for the majority of experiments, and these were: spatial frequency of 1°, Gaussian envelope σ parameter (orthogonal to carrier orientation) of 1.25°, aspect ratio 2:1 extended parallel to the carrier orientation, and drift rate of 18°/s. We adjusted these parameters if necessary to produce responses clearly above baseline when the stimulus was of 100% contrast but did not systematically search for optimal parameters. We attempted to keep the spatial dimensions small to avoid stimulus overlap while still adequately driving the cell. All stimuli moved in the preferred direction of the cell, as determined qualitatively during the original RF mapping.

Stimuli were presented in a rapid sequence with two frames between sequentially presented stimuli. Typically, a trial consisted of a sequence of 25 stimuli and was aborted if the monkey broke fixation at any point during the trial. Stimuli were presented in a horizontal array of five nonoverlapping locations, (see Fig. 1, bottom). We attempted to place the stimulus array over the RF so that at least one stimulus position was well-centered within the RF and at least one was near or at the edge of the RF. Within a single trial, presentations of single stimuli, pairs of stimuli at different locations, and blank intervals equivalent to an individual stimulus duration were interleaved.

The contrast of each stimulus was varied independently over a range of contrasts with approximately octave spacing. Typically, five or seven different contrasts, from 1 or 2% up to 64 or 100%, were used within an experiment. The location and contrast of each stimulus, or each member of a pair of stimuli, were pseudorandomly chosen. Each contrast was presented at each location alone and paired with all other contrasts. For the data in this paper, at least five stimulus repetitions were recorded for each possible pair-wise combination of location and contrast; for single-stimulus presentations, ≥20 stimulus repetitions were presented.

**Data analysis**

Times of spikes were corrected for the vertical location of the stimulus on the CRT screen and compiled into standard peristimulus time histograms (PSTHs). We calculated spike rates using a time window of 25- to 150-ms poststimulus onset for both single and paired stimulus conditions. We chose this time window based on a composite PSTH for all stimuli for all cells to avoid subjectivity in choosing response windows for individual cells. Additionally, this time window is the same as we previously used (see Fig. 3, Britten and Heuer 1999) in a related study addressing spatial summation in MT.

Firing rates for this time window were calculated and corrected for
maintained activity as estimated from interleaved “blanks.” These adjusted rates were then used for all subsequent analysis. All curve fitting was done using an iterative, maximum-likelihood method (STEPIT) (Chandler 1965). Likelihoods were directly estimated from the empirically measured experimental error.

RESULTS

The results from this experiment will be presented in two sections. First, we examine the effects of contrast at individual stimulus locations within the RF (1st-order properties). Then we address how space and contrast affect the interactions between pairs of stimuli (2nd-order properties).

Responses to single stimuli

Before we can discuss the interactions of multiple stimuli within the RF, we must first quantify responses to the individual stimuli at each location tested. We measured the response to single stimuli over a range of contrasts, approximately octave-spaced. The response to each contrast for a single cell represents an average of 20–80 stimulus repetitions. For each individual location within the RF, we found that the data were well fit using a hyperbolic ratio function of the form

\[ R(c) = \frac{R_{\text{max}}}{c^{n} + c_{50}^{n}} + M \] 

as shown for single positions for the example cell in Fig. 2. \( R(c) \) is the predicted response as a function of contrast \( (c) \), \( R_{\text{max}} \) is the maximum attainable response, \( c_{50} \) is the semi-saturation contrast, or the contrast at which half the maximum response is obtained, and \( n \) is an exponent that specifies the slope or steepness of the function. \( M \) represents the maintained activity and was fixed to the estimate provided by averaging activity over the interleaved blank stimulus periods. Previous studies have shown that the hyperbolic ratio function of Eq. 1 well fits the contrast responses of cortical cells in both V1 and MT (Albrecht and Hamilton 1982; Sclar et al. 1990). This was true in our data as well; the median fit captured 97.9% of the variance of the data, when each location was individually fit (see legend, Fig. 4, for description of the explained-variance calculation).

![Figure 2](image-url)

**FIG. 2.** Contrast response functions individually fit for each of the 5 stimulus positions within the RF. Symbols represent observed average responses and error bars depict SEs of the mean. Letters refer to stimulus locations, corresponding to Fig. 1B.

![Figure 3A](image-url)

**FIG. 3A.** Changes in \( R_{\text{max}} \) result in scaled replicas of the function.

![Figure 3B](image-url)

**FIG. 3B.** Varying \( c_{50} \) shifts the function to the right or left.

![Figure 3C](image-url)

**FIG. 3C.** Changing \( n \) alters the slope of the function, while leaving the maximum and half-max points unchanged.

However, it is clear by inspection that a single function will not account for all the stimulus locations—the data clearly do not lie along a single contrast-response function. To examine which response parameter(s) changed with respect to position, we performed a fitting procedure that allowed only a single designated parameter to vary across location. We produced fits for each individual location that differed solely in the value of a single parameter. Changes in \( R_{\text{max}} \) will result in scaled versions of the response function, as demonstrated in Fig. 3A. Figure 3B shows that allowing \( c_{50} \) only to vary produces a series of horizontally shifted functions that are otherwise iden-
tical. Allowing \( n \) to vary changes the slope of the function, as seen in Fig. 3C. We did not attempt to vary \( M \) as it represents the maintained activity of the neuron and was constant by definition.

Allowing either \( R_{\text{max}} \) or \( c_{50} \) to vary captured the data well, but allowing \( n \) to vary did not (data not shown). This was true for all the cells we recorded from. To assess which parameter accounted best for the changes in response with location, we calculated the percentage of variance explained by each, using a method described by Carandini et al. (1997). For the majority of cells, allowing \( R_{\text{max}} \) to vary captured more variance (median = 95.6%) than allowing \( c_{50} \) to vary (median = 94.5%), as shown in Fig. 4. Across our sample of 39 cells, this difference was significant (Wilcoxon paired-rank test, \( P < 0.001 \)), suggesting that the parameter that best captured spatial differences in stimulus effectiveness was \( R_{\text{max}} \). In turn, this indicates that contrast sensitivity does not vary across an MT cell’s RF, but response amplitude does. We chose to simplify our model of the single-stimulus responses from a set of five hyperbolic ratio functions (1 for each location, differing in \( R_{\text{max}} \)) to a single equation, with the spatial profile estimated by a Gaussian

\[
R = A\left(e^{(x - h)^2\sigma^2}} \right) c_{50} + M
\]  

(2)

In this expression, \( A \) is a scaling factor determining the maximum response, \( x \) is the stimulus location in degrees of visual angle, \( h \) is the center of the spatial Gaussian in degrees, and \( \sigma \) is the width of the Gaussian spatial profile. The other parameters are as described above. This model describes the data well, as seen for three example cells in Fig. 5, capturing on average 90.7% of the variance (range: 45.1–99.2%, median: 93.2%), and allows us to standardize the spatial location of the stimuli with respect to the size of the RF. This also allows us to derive a single semi-saturation contrast value for each cell. Analysis of the residuals from these fits showed that the modest loss of variance explained by simplifying the model in this way was not systematic. Therefore allowing fully independent response functions for each location was fitting noise in the data rather than systematic variation in the RF or contrast sensitivity.

**FIG. 4.** Comparison of explained variance by allowing either \( R_{\text{max}} \) or \( c_{50} \) to vary as a function of location. Percentage of explained variance is calculated by the expression: \( \% \text{var} = 100 \left[ 1 - \left( \frac{\text{model} - \text{data}}{\text{data variance}} \right) \right] \).
The distribution of $c_{50}$ values is shown in Fig. 6A. The median value across our sample of cells was 20.1%. This is higher than previously reported for MT cells (~7%) (Cheng et al. 1994; Sclar et al. 1990), consistent with the small size of the stimuli used here. Sclar et al. (1990) reported that $c_{50}$ varies inversely with stimulus area; their sample average of 7.6% was calculated using stimuli that filled the RF. Because spatial summation was proposed to explain the higher contrast sensitivity seen in MT compared with V1, it makes sense that in our experiments this difference should decline. The distribution of exponents is shown in Fig. 6B. The median exponent we observed was 3.57, very similar to the value of 3.0 previously reported (Sclar et al. 1990).

Responses to paired stimuli

Having established the effects of spatial location and stimulus contrast of individual stimuli for each cell, we can now turn our attention to the primary question of the effects of stimulus efficacy on interactions within the RF. Pairs of stimuli were presented interleaved with the individual stimuli described in the preceding text. Every combination of contrasts was tested for each pair of spatial locations within the receptive field.

Previously, we showed that responses to pairs of stimuli within the receptive field of MT neurons resembled a scaled version of linear summation (Britten and Heuer 1999). In these previous experiments, contrast was always 100%. In Fig. 7, we present similar analysis of the present data. In each panel, the $x$ axis represents the sum of the responses to the individual stimulus components of a stimulus pair, that is, the prediction of linear summation. The $y$ axis is the observed response to the paired stimulus presentations. If the responses to individual stimuli summed linearly, all points would fall along the unity diagonal (—). Perfect averaging of stimulus responses would fall along the dashed line, which has a slope of 0.5. We break the responses into three contrast categories: both stimuli of high contrast, both of low contrast, and mixed pairs where one stimulus was of high contrast and the other was of low contrast. The dividing line between high and low contrasts for this analysis was the $c_{50}$ from the single stimulus presentations. Summation varies in a sensible manner with stimulus contrast: where either or both members of the stimulus pair are of low contrast, then summation is approximately linear. This implies that the low-contrast member of the pair no longer is effective.
at normalizing the response to the higher contrast member of the pair. Consistent with our previous observations, when both members of the pair were of high contrast, all observations fell along a single line with a slope well below unity, indicating sub-linear summation. This cell is typical of MT cells, as shown in Fig. 8, where the same analysis is presented for the entire sample of MT cells.

This analysis pools both stimuli before relating this pooled quantity to a cell’s response, which might hide effects dependent on the relationship between the two individual stimulus contrasts. To investigate this, we pooled across the spatial locations of the stimuli and plotted the resulting average response (with each cell normalized to its own maximum rate) as a contour plot, shown in Fig. 9. This average response surface reveals two interesting features. The most conspicuous feature is the striking concavity visible over most of the surface, where either contrast is greater than ~30%. This concavity is particularly abrupt near either axis and reflects the loss of normalization from the stimulus of lower (near 0) contrast. Also evident in this figure is a convexity near the origin, where both contrasts are low. This is consistent with an expansive non-linearity when the total contrast is low.

This analysis, however, collapses across stimulus location, and ideally, we want to account for cells’ responses taking into account both the effects of location and of contrast. Our approach to this was to use descriptive modeling in an attempt to find the simplest (in terms of number of free parameters) model that would provide a good account of the main features of our data. All of the models we explored were loosely related to the divisive normalization model developed by Simoncelli and Heeger (1998). We have explored a family of related models and will present in detail the most successful one. The basic design of the model is a summation of first-order inputs, which we estimate as described in the preceding text, followed by a contrast-dependent normalization step. In this model, we allow the summation step (before normalization) a nonlinearity as well, as suggested by prior work (Britten and Heuer 1999). The form of the model is as follows

\[ R = (R_a + R_b)^{1/s} \left[ 1 - A \left( \frac{c_a}{c_a + \nu_{50}} \right) \left( \frac{c_b}{c_b + \nu_{50}} \right) \right] \]  

In this expression, \( R_a \) and \( R_b \) are the first-order responses to single stimuli. In this case, they are not the data values but instead are estimated from the contrast-dependent Gaussian model (Eq. 1), to reduce noise in the first-order estimates. The summation incorporates a nonlinearity, captured by the exponent, \( s \). We will take up the consequences of removing this nonlinearity in the discussion. Therefore the \( R \) terms incorporate both the first-order contrast dependence of the neuron and its spatial profile. The contrast dependent normalization is captured by the second term, which is dependent on the product of two hyperbolic ratio terms, capturing the contrast dependence of the normalization process. The \( \nu_{50} \) and \( z \) values in this expression capture the contrast dependence of the normalization, and \( A \) is an arbitrary scale factor to account for different degrees of normalization for different cells.

This analysis pools both stimuli before relating this pooled quantity to a cell’s response, which might hide effects dependent on the relationship between the two individual stimulus contrasts. To investigate this, we pooled across the spatial locations of the stimuli and plotted the resulting average response (with each cell normalized to its own maximum rate) as a contour plot, shown in Fig. 9. This average response surface reveals two interesting features. The most conspicuous feature is the striking concavity visible over most of the surface, where either contrast is greater than ~30%. This concavity is particularly abrupt near either axis and reflects the loss of normalization from the stimulus of lower (near 0) contrast. Also evident in this figure is a convexity near the origin, where both contrasts are low. This is consistent with an expansive non-linearity when the total contrast is low.

This analysis, however, collapses across stimulus location, and ideally, we want to account for cells’ responses taking into account both the effects of location and of contrast. Our approach to this was to use descriptive modeling in an attempt to find the simplest (in terms of number of free parameters) model that would provide a good account of the main features of our data. All of the models we explored were loosely related to the divisive normalization model developed by Simoncelli and Heeger (1998). We have explored a family of related models and will present in detail the most successful one. The basic design of the model is a summation of first-order inputs, which we estimate as described in the preceding text, followed by a contrast-dependent normalization step. In this model, we allow the summation step (before normalization) a nonlinearity as well, as suggested by prior work (Britten and Heuer 1999). The form of the model is as follows

\[ R = (R_a + R_b)^{1/s} \left[ 1 - A \left( \frac{c_a}{c_a + \nu_{50}} \right) \left( \frac{c_b}{c_b + \nu_{50}} \right) \right] \]  

In this expression, \( R_a \) and \( R_b \) are the first-order responses to single stimuli. In this case, they are not the data values but instead are estimated from the contrast-dependent Gaussian model (Eq. 1), to reduce noise in the first-order estimates. The summation incorporates a nonlinearity, captured by the exponent, \( s \). We will take up the consequences of removing this nonlinearity in the discussion. Therefore the \( R \) terms incorporate both the first-order contrast dependence of the neuron and its spatial profile. The contrast dependent normalization is captured by the second term, which is dependent on the product of two hyperbolic ratio terms, capturing the contrast dependence of the normalization process. The \( \nu_{50} \) and \( z \) values in this expression capture the contrast dependence of the normalization, and \( A \) is an arbitrary scale factor to account for different degrees of normalization for different cells.
This model provides a very good account of our data, as shown in Fig. 10. This figure depicts the pair responses as a function of the two stimulus contrasts for a single cell. For graphical clarity, the responses have been split into several groups, according to the $R_{\text{max}}$ of the first-order responses. Figure 10A shows the spatial profile of the RF in response to single stimuli of high contrast. Each stimulus location is labeled A-E; these location labels are referred to in the remaining panels to indicate which component stimuli are in each pair. Figure 10B shows a three-dimensional surface plot of the fits to the data for one stimulus configuration. The two locations are unequal in effectiveness as can be seen from the heights of the surface along each axis (where 1 or the other contrast is close to 0). Because it is a bit difficult to visualize the data with respect to the model surface, C–E show additional data for the same cell as families of two-dimensional plots.

In these plots, the contrast of one component forms the x axis, while different values of the second component form the different curves in each panel. In Fig. 10C, the responses are from locations C and D, near the center of the RF. Both locations produce strong and nearly equal responses to single stimuli. The lowest curve is where location D is at subthreshold contrast and thus shows the single-stimulus contrast-response function. As contrast is added to location D, the baseline (where location C contrast is subthreshold) rises systematically, as can be seen in the left portion of the plot. The uppermost curves, of course, come from cases where the contrast of location D is high, keeping the response high irrespec-
tive of the location C contrast. Note the dip in these curves at around 10% contrast. This dip reflects the onset of the normalization from location C. The fact that a dip exists demonstrates that the inhibitory effects of the contrast at this location appear at lower contrasts than do the excitatory effects. Also note the maximum value on this top curve does not rise much above the single stimulus (lower) curve; this is the primary consequence of the normalization. In $10D$, two modestly effective locations are used (note the change in vertical scale). Under these conditions, normalization is less complete, although it still engages at low contrasts. This is more obvious in the final panel, $10E$, where a completely ineffective stimulus location is paired with a central location (D in $10A$). In this case, all the top curves show responses to high-contrast stimuli in the effective location, and these responses are substantially attenuated once the location A stimulus reaches ~8–10% contrast. This shows that normalization becomes effective, and at low contrast, even when the stimulus is completely off the excitatory (“classical”) RF.

Our model captures all the main features of these data and some minor ones as well. In particular, a single normalization weight, independent of spatial location, and a single contrast dependence are sufficient to explain the manner in which the curves vary with different stimulus locations. For this cell, this model captures 87% of the variance in the data. Across the population, this model explained a median of 78.6% of cell response variance. Inspection of residuals to the fits did not show a systematic deviation as a result of contrast, spatial location, or overall response across cells.

The main question being addressed by this model was the quantitative dependence of normalization on contrast. Three parameters from the model—$A$, $v_{50}$, and $z$ in expression 3—capture different aspects of the contrast-dependent response normalization, and the distributions of the best-fit values of these parameters are given in Fig. 11. The scale parameter, $A$ (Fig. 11A) varies around a median of 0.29. This scale parameter constrains the maximal amount of normalization; the effects of changing $A$ are described in the following text and one example is graphically illustrated in Fig. 12D. The saturation constant for normalization, $v_{50}$, is the contrast value where normalization becomes half-maximally effective and is directly comparable to the $c_{50}$ term that describes contrast responses for single stimuli. The median value for this term, whose distribution is shown in Fig. 11B, is 14.8%, which is a noticeably lower value than for $c_{50}$. Therefore normalization becomes effective at quite low contrasts, where excitatory responses are still below their half-maximal value. Furthermore, the values of $v_{50}$ and $c_{50}$ are not significantly correlated ($r = -0.18, P > 0.2$), suggesting that the contrast sensitivity of the normalization does not directly arise from the sensitivity of the excitatory processes driving the cell. Last, the distribution of values for the normalization exponent, $z$, which characterizes the steepness of the normalization as a function of contrast, is shown in Fig. 11C.

Each of these parameters affects the normalization in slightly different ways. To illustrate this graphically, we show the effects of parameter changes on the predicted pair response in Fig. 12. Each plot describes the response as a function of the two stimulus contrasts, plotted as in Fig. 10, C–E. For simplicity, we have only portrayed responses at two equally effective stimulus locations. Figure 12A shows the responses generated by a model exemplifying the median values from our MT fits. Increasing $v_{50}$, as shown in Fig. 12B, increases the contrast at which the normalization takes effect; here we’ve raised it from the median value of ~15 to 25%. This change eliminates the “dip” in the responses—normalization does not become effective until the excitatory response from the second component is adding substantially to the output. Changing $z$ alters the rate of increase of normalization, which creates a deeper “dip” in the responses, as seen in Fig. 12C. Figure 12D illustrates the effects of altering $A$. For this panel, we’ve lowered $A$ to a value of 0.16. This reduces the impact of the contrast-dependent normalization, causing two changes in the responses. First, the dip disappears because normalization no longer can suppress the increasing excitation caused by the second component. Second, the maximum response when both

![Figure 11](http://jn.physiology.org/DownloadedFrom/10.22032/jn-88-6-2002-12-30.pdf)
DISCUSSION

In this paper, we explored the effects on summation of varying the contrasts and locations of multiple stimuli presented within single MT RFs. The main results were twofold. First, when single stimuli were presented at different locations, the responses could be best described as a single invariant contrast-response function, scaled differently at different locations within the RF. Second, the interaction between stimuli (divisive normalization) was very sensitive to stimulus contrast. This divisive normalization was near saturation by the time that both of the stimuli were of contrast greater than \( \sim 15\% \). In this discussion, we will relate these observations to previous work and consider their implications for the mechanisms underlying divisive normalization in extrastriate cortex.

Relationship to previous work

Previous work from this and other laboratories has documented that responses in extrastriate cortex to multiple stimuli are less than that expected on the basis of linear summation (Britten and Heuer 1999; Recanzone et al. 1997; Snowden et al. 1991; Treue et al. 2000). The present work confirms and extends these findings. In previous work from our own laboratory, single- and multiple-stimulus experiments were not randomly interleaved, raising the possibility that contrast adaptation (a slow process) might have contributed to the results. The present experiment, which contains an internal replication of this work, produced identical results for the overlapping conditions (data not shown), ruling out this interpretation.

Another conclusion from the previous experiment was that normalization was effective for stimuli placed at the fringes of the classical RF of the MT cells. The present data support this conclusion. In our descriptive modeling, we account for the space dependence of the responses only in the numerator of the divisive normalization step; this captures the classical RF of the cell. The denominator of the model depends only on contrast and not on location yet describes our data well. This again shows that contrast anywhere in the vicinity of an MT cell’s RF is equally effective at normalization.

One previous experiment has investigated the contrast sensitivity of MT cells (Sclar et al. 1990), and it is useful to compare the two experiments. In the study of Sclar et al., contrast sensitivity of MT cells to centered gratings was measured. This was found to be quite high, higher even than that of magnocellular neurons in the LGN (which provide the bulk of input to MT) (see Nealey and Maunsell 1994). The authors concluded reasonably that the size of MT RFs allowed for spatial summation of contrast, increasing contrast sensitivity. At face value, our data are very consistent with this interpretation. The contrast sensitivity of MT cells measured to our single stimuli (which approximate the size of a V1 RF) was substantially lower than that observed by Sclar et al. (our estimate of \( c_{50} \) was \( \sim 20\% \); theirs was 7.6%). To investigate whether the lower sensitivity of our sample was due to stimulus size, we performed an analysis on a subset of the data. For each cell, we calculated contrast response functions for each pair of stimulus locations when both stimuli were of equal contrast. We fit these data, allowing \( R_{\text{max}} \) to vary across pair location, and compared the resulting unique \( c_{50} \) values generated by fitting the single stimulus locations with \( R_{\text{max}} \) free to vary. When two stimuli of equal contrast are presented, we estimate

stimuli are at high contrast is noticeably higher. This would in turn result in contrast-dependent responses that would presumably be nonoptimal from a coding standpoint. In any case, this analysis illustrates how the summation surface is shaped by a delicate balance between excitatory and inhibitory influences.

Figure 12. Effects of varying different parameters in the normalization model. Contrast for 1 stimulus location increases along the x axis; each curve represents a single contrast level for the other stimulus location. A: reference curves, which have parameter values set to values typical for our sample. B: higher \( \nu_{50} \)—note that the dip in the high contrast curves caused by normalization is flattened as it occurs at higher contrasts. C: higher \( z \)—note the dramatically steeper inflection of the high contrast curves. D: lower \( \nu \)—note the inflection is almost eliminated and the overall response is higher, indicating a reduction in normalization.
a slightly, but significantly, lower $c_{50}$ of 18.5% (Wilcoxon paired rank test, $P < 0.02$). This reduction in $c_{50}$ is to be expected from summation across space and is generally consistent with the observations of Sclar et al. (1990).

Robustness of the model

Our model of normalization was a little unusual, and we wanted to test some of its assumptions. Overall, it contained nine free parameters: five to describe the first-order responses and four describing the second-order interactions. We were particularly concerned about the summation nonlinearity ($s$ in Eq. 3) in the numerator. This parameter modestly improved the overall quality of the fits as it did when contrast was not varied (Britten and Heuer 1999). The average improvement in percentage of variance explained was 1.8%; this was a significant improvement in the majority of cases (70%; nested likelihood ratio test, $P < 0.05$). This term contributes most heavily to the fit where the response amplitudes are high (near the center of the RF, stimuli of high contrast) in a regime where the contrast normalization is effectively saturated. Furthermore, it doesn’t interact in any important way with the main point of the model—estimating the contrast dependence of normalization. When this term is removed from the model, the contrast-dependent normalization terms change $< 10\%$.

Another unusual feature of our model is the multiplicative interaction between the two contrasts in the denominator. In a more conventional normalization model (e.g., Simoncelli and Heeger 1998), the quantity responsible for normalizing responses is dependent on the sum of local contrast-dependent responses, not the product. We explored this class of models first before exploring the more successful model we have implemented in this paper. We fit our data with the model of Simoncelli and Heeger that requires an additional parameter to stabilize the denominator when both contrasts are low. Despite this additional parameter (which should in principle allow the model to perform better), the additive model generally produced poorer fits to the data. The multiplicative model fit better in 75% of the cells, and the median improvement in percentage of variance explained was 1.2%. As described in RESULTS, inspection of the fits suggested the mechanism: the multiplicative model allowed rapid release from normalization as either stimulus contrast neared zero. However, it is in general clear that both additive and multiplicative forms of the normalization fit the data well, so further experiments targeting this question are clearly required. We emphasize the multiplicative version because it worked better for our data and because it is a possibility not, to our knowledge, previously considered.

Mechanisms of normalization

It is not in question that in most cases responses to multiple stimuli are not as great as that expected by summation. However, the mechanism for this phenomenon remains a matter of some dispute. Despite their recent popularity, divisive normalization models are not the only candidates. Many of the phenomena attributable to recurrent, divisive scaling can also be explained by synaptic depression (Abbott et al. 1997). We believe that the present observations exclude such single-cell mechanisms. Synaptic depression, where single synaptic inputs become weaker due to repeated use, is a viable candidate for contrast gain control where untuned input (e.g., LGN cells) impinging on tuned cells (orientation-selective cells in striate cortex). In MT, there is good evidence that spatial pooling groups inputs that are already tuned for direction (Movshon and Newsome 1996). Thus the different stimuli in the present experiments are activating distinct inputs. In the cases where our stimuli are maximally distant, these inputs are probably effectively nonoverlapping. In this case, one stimulus will not be capable of adapting the synapses responsible for the response to the other stimulus, and a simple feed-forward depression model will fail. However, recurrent models are completely consistent with the present observations.

The contrast dependence of the normalization that we have measured also helps to shed some light on the normalization mechanism. The fact that normalization is engaged at very low contrasts, where the response of even the most active elements is still low, also helps to reject afferent synaptic depression for a mechanism. Furthermore, the success of a model that incorporates a multiplicative term between the different contrast-dependent elements further suggests a local network basis for the normalization. While single cells appear to multiplicatively combine their excitatory inputs in some cases (Pena and Kohnishi 2001), it seems unlikely that a single feed-forward connection could both multiplicatively combine signals and also divide a cell’s output by the product. On the other hand, local circuit connections within and across cortical columns contain both excitation and inhibition in abundance. Such local, diffuse connections might endow the column with both the cooperative and divisive aspects that our experiments reveal. It seems likely that intracellular recording and local circuit tracing will be necessary to test this hypothesis. Work of this sort has been little done in MT before now but clearly would be highly useful.

The foregoing suggests that circuits local to MT might carry out the normalization, but it seems equally likely that recurrent signals from other cortical areas are also involved. Both feedforward and feedback connections originate from pyramidal cells, and are thus likely to be excitatory (Jones and Wise 1977; Maunsell and Van Essen 1983a). This recurrent excitation could provide the cooperativity that our results suggest, and the divisive component might result from such feedback connections impinging on inhibitory interneurons in MT. Obviously, this architecture would be spatially coarse-grained, which is also consistent with our results. Of course, the local circuit and inter-area feedback hypotheses are not exclusive, and the truth is likely to embody some of both.

The authors thank R. E. Tarbet, J. L. Moore, M. R. Nilsson, and H. R. Engelhardt for technical assistance and support. The display software was written by A. Jones. We also thank D. Heeger, P. I. Harness, and T. Zhang for useful discussion.

This work was supported by National Eye Institute Grant EY-10562 to K. H. Britten and by Vision Core Grant EY-12576.

REFERENCES


J Neurophysiol • Vol 88 • December 2002 • www.jn.org