Responses of Primate Visual Cortical V4 Neurons to Simultaneously Presented Stimuli

TIMOTHY J. GAWNE AND JULIE M. MARTIN
University of Alabama, Department of Physiological Optics, Birmingham, Alabama 35294

Received 22 February 2002; accepted in final form 14 May 2002

INTRODUCTION

Visual processing in cortex is often considered to be an at least partially hierarchical building up of larger receptive fields from smaller ones. For some neurons, the building up of a large receptive field is well modeled as an approximately linear combination of signals from neurons with smaller receptive fields, most notably the simple cells of visual cortical area V1. Nevertheless, it seems unlikely that linear summation is the only operator that a neuron could apply to its inputs. It has been proposed (primarily on theoretical grounds) that one particularly useful operator would be the “MAX” operator, that is, a neuron’s output is determined by the input with the maximum firing rate. (Riesenhuber and Poggio 1999) Very roughly, a MAX operator could in effect select from a large number of inputs without explicit scanning, an operation that could play a particularly important role in helping to generate neuronal responses that are invariant over some parameters, such as the position or size of a visual shape. However, the existing literature gives relatively little experimental evidence in support of the MAX model. Sato explored the interactions of two bar stimuli in the receptive fields of inferior temporal neurons and found that the responses to two stimuli simultaneously were often well fit by the maximum of the response to each separately (Sato 1989). The lack of additional evidence for the MAX model is most simply explained by the lack of studies that explicitly test this hypothesis.

Gawne, Timothy J., and Julie M. Martin. Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. J Neurophysiol 88: 1128–1135, 2002; 10.1152/jn.00131.2002. We report here results from 45 primate V4 visual cortical neurons to the preattentive presentations of seven different patterns located in two separate areas of the same receptive field and to combinations of the patterns in the two locations. For many neurons, we could not determine any clear relationship for the responses to two simultaneous stimuli. However, for a substantial fraction of the neurons we found that the firing rate was well modeled as the maximum firing rate of each stimulus presented separately. It has previously been proposed that taking the maximum of the inputs (“MAX” operator) could be a useful operation for neurons in visual cortex, although there has until now been little direct physiological evidence for this hypothesis. Our results here provide direct support for the hypothesis that the MAX operator plays a significant (although certainly not exclusive) role in generating the receptive field properties of visual cortical neurons.

METHODS

Recordings were made from 45 single V4 neurons in two awake rhesus monkeys (Macaca mulatta). The monkeys were anesthetized with isoflurane, and a stainless steel recording chamber was implanted over the dorsal posterior surface of the skull. A coil of Teflon-coated stainless steel wire (Cooner wire AS632) was inserted under Tenon’s capsule of one eye (Judge et al. 1980) allowing eye position to be monitored via the magnetic field/search coil technique (Riverbend Instruments) (Robinson 1963).

After recovery the animals were trained to fixate on a small spot of light on a computer monitor. The animals were given juice rewards for maintaining eye position to within ±½ a degree of fixation, although in practice their actual performance was typically much better (average deviation from fixation: ± 0.5° of arc). The video display was run at 85 Hz (Colorgraphics), positioned 59 cm away from the eye, and was 39 cm wide by 28 cm tall. To provide sufficient room to position, the stimuli the fixation point was typically moved to a far corner of the display, depending on the receptive field location of the neuron under study.

Single-unit recording was made by advancing polymide-coated tungsten microelectrodes of approximately 1.2 MΩ impedance (Microprobe) through a 23-gauge stainless-steel guide tube that penetrated the dura. In one animal, the guide tube was positioned to allow recording from ventral V4; in the other animal, recordings were made from dorsal V4.

The stimuli were high-contrast black-and-white patterns presented on a uniform gray background. There were seven stimuli, each selected from the set of two-dimensional Walsh patterns (Kjaer et al. 1997). These stimuli were chosen because they are easy to generate and have a large range of spatial frequency and edge variation, and experience has shown that at least one of these patterns can elicit a strong response from a significant fraction of the neurons in many visual cortical areas. This study was done in combination with another more extensive one. When a well-isolated single neuron was acquired, it was tested with this specified stimulus set, and if no strong responses were elicited, the neuron was bypassed for this study: no
attempt was made to find an optimal set of stimuli for each neuron. Neurons with receptive fields that were not large enough to easily encompass two clearly separated stimuli were also bypassed. We placed the two stimulus locations such that they were clearly separated and at least one stimulus gave a similarly strong response at each location. The stimuli were sized only to ensure that they could both fit into a receptive field at one time and that a broad range of response strengths could be elicited. No further attempts to optimize stimulus location were made. We required that at least one stimulus at each location elicit a response at least twice that of background (statistically significant t-test \( P < 0.01 \)) and that the range of firing rates for all seven patterns be at least 2:1 and also significantly different by t-test \( (P < 0.01) \). The neurons in this study are 45 of 163 isolated neurons found in the course of other studies.

The stimuli were shown separately in two locations and, for all 49 combinations of simultaneous presentations at the two locations, in shuffled random order a minimum of 10 trials each (mean: 15.2 trials, maximum: 24 trials). The stimuli were flashed on for 24 video frames (at 85 Hz approximately 282 ms duration) with the eyes fixed. This stimulus duration was chosen because it is similar to the normal primate visual experience of three to four periods of stable vision per second punctuated by rapid saccadic eye movements. The results of studies in both V1 and inferotemporal visual cortex have demonstrated that static flashed stimuli generally produce neuronal responses that are similar to those elicited by saccadic eye movements (DiCarlo and Maunsell 2000; Richmond et al. 1999). Trials where the eye moved were discarded. The background luminance was 33 cd/m², and the stimuli had an average luminance matched to the background with a Michelson contrast between the black-and-white stimulus regions of 78%.

Action potentials from single units were identified off-line, and unit times were determined to 1 ms of precision. Only clearly isolated units with an absolute refractory period were counted as single units. Spike density functions were created from the individual spike times by convolving with a \( \sigma = 3 \) ms Gaussian kernel and then averaging (Heller et al. 1995; Silverman 1986). Mean firing rates were calculated by averaging the number of action potentials in an interval starting 20 ms after stimulus onset (before the earliest response) and lasting 200 ms in duration (ending before any off-response). Response latencies were calculated as the point at which the spike density reached one-half of the peak value (Gawne et al. 1996).

The ability of the MAX model to fit the data was compared with two other models: linear summation and weighted average. Linear summation was chosen as one alternative because it is easy to imagine how a neuron integrating synaptic inputs might act as an approximately linear summator and because many neurons in the visual system have been demonstrated to operate in a quasi-linear manner. Weighted average was chosen as the other alternative because it has already been proposed that this is how neurons in area V4 combine inputs from different subregions of their receptive fields and because of its proposed role in models of biased competition in attention (Reynolds et al. 1999). Because the issues involved in discriminating between these models are qualitatively different, these comparisons were done separately.

Determination of cortical site was made by a combination of comparing stereotaxic coordinates with a standard atlas (Paxinos et al. 2000), comparing receptive field location and size, and in one animal (ventral V4; superior visual fields) by making 40-μm coronal sections and Nissl staining; several sites were marked with fluorescent microspheres for reference.

All experimental procedures and care of the animals were carried out in compliance with guidelines established by the National Institute of Health and were approved by the University of Alabama at Birmingham Animal Care and Use Committee.

**RESULTS**

Recordings were made from 45 well-isolated single neurons. The sizes and locations of the two stimuli used for each neuron are illustrated in Fig. 1 (only 43 pairs are shown here because for two cases the same stimulus configuration was used twice). The stimuli elicited comparable responses at both locations; taking the strongest response for a single stimulus at either location, the strongest response for a single stimulus at the other location was on average 86.3 ± 12.7% (mean ± SD) of the maximum response at the other location (in Fig. 2B, the maximum response to a stimulus at the location represented by the top row was 66.7% of the maximum response to a stimulus at the location represented by the leftmost column; this was the worst case).

For many neurons, we could not determine any clear pattern to the simultaneous presentation of two stimuli. However, there were a significant number of neurons for which the MAX model appeared to fit the data fairly well. Figure 2 shows two examples of this: the responses to the simultaneous presentation of two stimuli are generally closest to that of the stimuli that gave the largest response when presented in isolation. Figure 3 shows plots from the two example neurons in Fig. 2. Here the actual response magnitude is plotted on the horizontal axis and the prediction made by the MAX model (maximum of the responses to each stimulus separately) and the prediction made by a linear model (response to both combined equals the sum of the responses to each separately) are plotted on the vertical axis. While not a perfect fit, the MAX model predicts the responses of the neurons to two simultaneous stimuli fairly
well, much better than does a linear model. Because this relationship holds even for cases where the stimuli do not elicit maximal responses, the success of the MAX model cannot be due to saturation.

Figure 4 A shows the residual mean-squared error (normalized as a fraction of the total power in the neuronal responses) for all 45 neurons in this study, comparing a MAX to a linear model. For most of the neurons in this study, the MAX model does a much better job of accounting for the variance in the responses than does linear summation. For 24/45 neurons a MAX model accounted for more than 90% of the power (sum of squares) in the responses.
We developed another metric to more rigorously test the ability of the MAX model to predict neuronal responses. If the response to one or both stimuli presented separately is near the maximum for that neuron, then the success of the MAX model for the simultaneous presentation of both could be due to saturation. Also, if the response to one stimulus is near zero, then it is perhaps not surprising that the response to both stimuli combined is close to the response of the single stimulus that alone elicited a significant response. The condition where each stimulus separately gives a response that is one-half-maximum should provide a particularly stringent test of the MAX model. Therefore we calculated the mean absolute value of the residual (normalized to a fraction of full-scale) of both the linear and the MAX models, across all stimuli, but with a weighting function for each stimulus combination defined as

\[ w_{ij} = \frac{1}{2}(1 - \frac{|R_i/R_{MAX}|}{0.5})(1 - \frac{|R_j/R_{MAX}|}{0.5}) \]

where \( R_i \) is the response to stimulus \( i \) at one location alone, \( R_j \) is the response to stimulus \( j \) at the second location alone, \( R_{MAX} \) is the maximum response to any single stimulus alone, and \( w_{ij} \) is the weight applied to each condition of two simultaneous stimuli.

The results of using this weighting function are shown in Fig. 4B. Even with this more stringent test the MAX model does much better than the linear model for most of the neurons studied here.

We also explored the degree to which our data could be explained by the weighted-average model of Reynolds et al. (1999), in which the response of a V4 neuron to two stimuli is modeled as the weighted average of the responses to each stimulus separately. In general we did not find that the weighted average model fit our data as well as the MAX model, although there are some subtleties to comparing the two

---

**Fig. 3.** The actual mean firing rate (horizontal axis) vs. the predicted firing rate (vertical axis) for all 49 cases where 2 stimuli were presented simultaneously for the 2 example neurons shown in Fig. 2 (A and B). ○, the prediction made by a linear pooling model; +, the prediction made by the MAX model.

**Fig. 4.** A: the residual mean squared error (MSE) of the MAX model (vertical axis) vs. the residual MSE of the linear model (horizontal axis), normalized to the total power in the signal. B: the residual error is normalized as a fraction of the maximum responses and weighted to emphasize conditions where both stimuli give responses equal to one-half-maximum (see text). The points A and B within each panel refer to the specific examples in Figs. 2 and 3.
models. The weighted average model requires one free parameter. We forced the value of the weight to lie between 0.4 and 0.6 (the range typically found by Reynolds et al.), this is because there were conditions (such as Fig. 2B) where one stimulus tended to produce slightly less strong responses than that of the other stimulus, and in this case the weights would otherwise adjust to 1.0 and 0.0, ignoring the weaker stimulus entirely and converting the weighted average into a de-facto MAX operator.

Whenever two stimuli each elicit responses that are of comparable magnitude the weighted average model and the MAX model make similar predictions, thus it was a general rule that when the MAX model did a relatively good job of fitting the data the weighted average model also did relatively well, and the overall differences in goodness of fit between the MAX and weighted mean models are smaller than between the MAX model and linear summation. Using the criterion used in Fig. 4A, the residual mean-squared error for the MAX model correlated with the residual mean-squared error for the weighted average model \( r = 0.83 \) (see Fig. 5). There were not two separate populations of neurons, one of which was well described by the MAX operator and one of which was well described by the weighted average, rather both models tended to do relatively better or relatively worse at fitting the data for the same group of cells with the MAX model typically doing better than the weighted average model when the fits were good and both models performing on average similarly when the predictions were poor. In general the weighted average only did better than the MAX model for cases where the fits for both were poor.

The best way to discriminate between the MAX and the weighted average is to focus on the condition where one stimulus by itself generates a maximal response, and the other stimulus by itself generates a response that is less than half as great. On average, the response of a neuron to an optimal stimulus at one position is increased by \( 5.1 \pm 1.5\% \) by the presence of a nonoptimal (response less than half that of the optimal) stimulus at the other location. This is a slight tendency toward linear summation, but clearly the data are still closer to a MAX model than to linear summation or else the MAX model would not have fit the data better than the linear one. There are examples where the presence of a nonoptimal stimulus reduced the firing of a cell that was responding to the optimal stimulus, but these are infrequent. In general, the addition of a weak stimulus to a strong stimulus does not consistently reduce the firing rate of the cell, as would be expected for a weighted average model, indeed on average there seems to be a slight enhancement.

Given the angles and depths of our recording, we are not able to state with confidence what cortical layers we had recorded from. It might have seemed reasonable that there would be a positive correlation between the degree to which the MAX model fit the data and the degree of separation of the stimuli, but this was not the case. The correlation between the degree to which the MAX model fit the data, and the separation of the two stimuli in degrees was only \( r = 0.14 \), not statistically significant. However, given that the stimuli were positioned at the maximum degree of separation within each neuron’s receptive field, this is perhaps not so surprising, as the proportional degree of separation within the receptive field was relatively constant.

We were not able to rigorously test the degree to which the simultaneous presentation of two stimuli affected the temporal pattern of the responses. This is because even with data compression techniques, it generally requires three parameters to characterize the temporal variation of a visual cortical neuron’s responses (Heller et al. 1995). With between two and six individual stimuli per neuron that elicited a strong response, there is insufficient data to reliably characterize a response in three dimensions. Thus we cannot perform a full analysis of the interactions in the temporal domain of the responses to different combinations of stimuli. Nevertheless, it was the case that the response to two stimuli presented simultaneously generally followed the time course of the strongest response to each stimulus presented separately. To quantify this, we calculated the correlation coefficient between the spike density function of the responses to both stimuli presented at the same time, with the spike density functions to the strongest response to each separately, across all combinations of stimuli (Fig. 6). For neurons where the MAX model fits the data moderately well, the correlation between the waveforms typically ranges from between \( r = 0.71 \) and \( r = 0.94 \).

In particular, we found that when a stimulus that elicited a short-latency response was combined with a stimulus that elicited a long-latency response, the temporal pattern of the response to both combined was not a mixture of the two but rather followed the pattern set by the stimulus eliciting the shortest latency (see Fig. 7). In most cases, the shorter-latency response was at all time points either equal or stronger than the longer-latency response, so we can only show isolated examples of a short-latency response suppressing a longer-latency response that would have been stronger at a later point in time. Nevertheless the effect was striking when the conditions to see it were present. This suggests some form of temporal gating mechanism for the MAX effect, but as we saw the MAX effect equally as well for two stimuli that each elicited responses of comparable strength and latency, temporal gating cannot be the entire explanation for this phenomenon.

More generally, we also found that the response latency for
two stimuli combined was locked to the latency of the shortest response (see Fig. 8): the presence of a stimulus that by itself elicits a relatively longer-latency response neither advances nor retards the response latency to a stimulus that by itself elicits the strongest response, especially for neurons for which the MAX model provided a good fit (leftmost portion of the graph).

DISCUSSION

This study provides direct physiological evidence that the MAX operator plays an important role in generating the properties of the receptive fields of visual cortical neurons, consistent with previous suggestions that this should be so (Riesenhuber and Poggio 1999).

We make no claims as to the relative frequency of neurons in visual cortical area V4 for which the MAX model outperforms a linear model; given the selection criteria of our study all that we can say is that there are many neurons in V4 for which a MAX model does a creditable job of predicting the response to the simultaneous presentation of two spatially separated stimuli. We could not determine any clear pattern to the responses of the neurons for which the MAX model did not fit: presumably there are more complex nonlinear operations being performed in addition to the MAX operator.

Reynolds et al. explored the interactions between a single reference bar and 16 probe bars varying in color and orientation in 18 V4 neurons (Reynolds et al. 1999). They concluded that the responses to two stimuli presented simultaneously are a weighted average of the responses to the stimuli presented separately. A similar finding was also found in another study by this group with a more limited stimulus set (Kozloski et al. 2001). Their finding that for two stimuli that elicit equally large responses the response to both together is similar to either alone is consistent with our findings. However, their finding that for two stimuli that separately elicit a strong and a weak response the response to both combined is midway between the two is not something that we consistently found.

However, Reynolds et al. did find some neurons where the response to a strong stimulus was essentially unaffected by the simultaneous presence of a weaker stimulus; thus our results lie within the range of their data. Conceivably the difference between our results and theirs is due to the different stimuli used (bars vs. Walsh patterns), however, as different classes of stimuli seem to elicit qualitatively similar patterns of response in visual cortex (Wiener et al. 2001), it seems unlikely the differences between bars and Walsh patterns could explain our results. We feel that the most likely explanation for our inability to find a consistent weighted-average effect is due to the degree of separation of the stimuli in the two studies. In this study, we specifically selected neurons with large receptive fields that allowed two square stimuli to easily fit inside with a large degree of separation; conceivably this may have increased the likelihood of finding evidence for the MAX operator by limiting interactions at earlier processing stages. While
Reynolds et al. took care to ensure that their stimuli did not overlap, they made no effort other than this to increase the separation, and thus we feel that the weighted average effect is most likely due to some combination of contrast gain control and surround suppression effects arising earlier in the visual system. Our results here in no way take away from the classic and surround suppression effects arising earlier in the visual system that have smaller receptive fields. If these fields were constructed only by linearly summing inputs from smaller receptive fields, the result would only be a progressive blurring: there must be nonlinear operations as well. It is reasonable to hypothesize that there are only a limited number of primitive classes of operations that neurons use to combine signals from neurons with smaller receptive fields. Some operations may be highly complex, possibly involving feedback loops and hard to analyze with simple experiments. Other operations may prove easier to uncover. Linear summation has been identified as one such operation, for example in the simple cells of primary visual cortex. The MAX model from limited data sets. Unlike the linear model, taking the maximum of two noisy samples is inherently biased upward. Consider an extreme case: if there were 100 separate stimulus locations, and each stimulus by itself gave a response of identical magnitude, the maximum of 100 separate noisy measurements would likely be unusually high and thus would tend to make the prediction of a MAX model higher than it should be. Because we only used two stimuli, this effect was relatively small for this study. Using the maximum of the responses minus one-half of the standard error of the mean to predict the MAX value (an approximate compensation for the upward bias of a MAX operation) resulted in no significant changes in the results. Nevertheless, it must be born in mind that when evaluating nonlinear models whose parameters are based on noisy estimates there can be subtle bias effects that must be watched for. In the case of the MAX model, these effects will be strongest when there are only a few trials per stimulus condition combined with large numbers of simultaneously presented stimuli. These effects are in addition to the more obvious problem that with 63 separate noncontrol conditions per neuron the odds are that there will be at least a few unusually strong or weak responses.

This study was done in a purely preattentive paradigm. Based on the results of other studies in V4 it is likely that using a paradigm where the animal had to pay selective attention to one specific location within the receptive field would have changed the response patterns (Luck et al. 1997; Reynolds et al. 1999). It might be argued that a MAX operation could be mimicked by selective attention, i.e., by the focus of attention shifting to the most salient stimulus. However, by using short stimulus durations and by showing that the MAX operation is in evidence in even the earliest part of the response, this does not seem like a very likely explanation.

The relationship between the magnitude of a neuron’s response and the latency is not necessarily fixed, but may vary according to the parameters of a visual stimulus (Gawne et al. 1996). However, all other things held constant, it is generally the case that stronger responses are associated with shorter latencies and weaker responses are associated with longer latencies. For a linear model, the response to two stimuli combined should be larger than that of either alone and thus potentially of shorter latency. For the weighted average model, the response to two stimuli that elicit unequal responses should be smaller than the largest response and thus potentially of longer latency. Thus to the extent that response latency covaries with magnitude, the results here are most consistent with the MAX model. More generally, the finding that the temporal pattern of the response to two stimuli is locked to the pattern of the strongest response to an individual stimulus fits with the proposed functional role of the MAX operator as performing a selection of inputs; one input is effectively selected out of many with minimal effect of the presence of other stimuli.

![Figure 8](http://jn.physiology.org/)

**Figure 8.** Histograms of the response latency to both stimuli at once, centered about the time of the shortest response latency of each stimulus shown separately. *A:* results across all data in the study for which a reliable latency could be determined. *B:* results limited to the 12 neurons for which the MAX model fit the data best (mean-squared error). *C:* results further limited to those in which the differences in the response latencies to the stimuli presented separately exceeded 10 ms. The response latency to 2 stimuli at once was set by the shortest response latency to each stimulus separately; there is no consistent advance or delay of the response onset by the presence of a stimulus that by itself elicits a shorter latency response.
operator would appear to be another such primitive operator, and as such, we would hypothesize that it will be found in multiple areas of the visual system.

We thank M. Bolding, A. Yildirim, and J. Millican for technical assistance, A. Dobbins for general comments, R. Weller for assistance with the anatomy, and D. Gawne for editorial assistance.

This work was supported by McDonnell-Pew Program in Cognitive Neuroscience Grant 96-27 and National Eye Institute Grant EY-11552-01 (both to T. Gawne).

REFERENCES


