A multiplicative model for spatial interaction in the human visual cortex

Xian Zhang
Department of Psychology, Columbia University, New York, NY, USA

Jason C. Park
Department of Psychology, Columbia University, New York, NY, USA

Jennifer Salant
Department of Psychology, Columbia University, New York, NY, USA

Sonya Thomas
Department of Psychology, Columbia University, New York, NY, USA

Joy Hirsch
Functional MRI center, Department of Radiology, Columbia University, New York, NY, USA

Donald C. Hood
Department of Psychology, Columbia University, New York, NY, USA

Multifocal visual evoked potentials (mfVEP) were recorded simultaneously for both the target and the neighbor stimuli, each varying over 6 levels of contrast: 0%, 4%, 8%, 16%, 32%, and 64%. For most conditions, the relationship between the amplitude of target response and the contrast of the neighbor stimulus, as well as the amplitude of the response to the target stimulus, were described with a simple, normalization model. However, when the neighbor stimulus had a much higher contrast than the target stimulus, the amplitude of the target response was larger than the prediction from the normalization model. These results suggest that spatial interaction observed in the mfVEP requires (1) multiplicative mechanisms, (2) mutual inhibition between neighboring regions, and (3) a mechanism that saturates when the ratio between the contrasts of the target and that of the neighbor is large. A modified multiplicative model that incorporates these elements describes the results.

Keywords: contrast response function, spatial interaction, normalization model, V1, multiplicative model


Introduction

Spatial interaction is a term that refers to the influencing of the response to a target stimulus by an adjacent stimulus. Although spatial interaction has been demonstrated in numerous studies of the visual cortex, the mechanism(s) of this interaction is still unclear. A simple model often does not explain the data from previous studies in spatial interaction.

One typical approach for studying spatial interaction involves varying the contrast of the target stimulus while the contrast of the neighbor stimulus is set at several levels, in other words, measuring the spatial interaction effect on the contrast response function (CRF) (i.e., the relation between the stimulus contrast and the response magnitude). The CRF of the visual system is nonlinear. For example, the CRF of most V1 neurons has been satisfactorily described with a nonlinear equation (Albrecht, Geisler, Frazor, & Crane, 2002; Naka & Rushton, 1966; Sclar, Maunsell, & Lennie, 1990),

\[ R = \frac{R_{\text{max}} C_t^\alpha}{\sigma^\alpha + C_t^\alpha}, \]

where \( R \) is the amplitude of the response to stimulus \( t \), the target stimulus, \( R_{\text{max}} \) is the asymptotic amplitude of the response, \( C_t \) is the contrast of the stimulus, \( \alpha \) is the exponential term that alters the steepness of the CRF, and \( \sigma \) is the semi-saturation contrast. Although this equation is only descriptive, it is thought that the nonlinearity may be due to the interactions among the neurons responding to the stimulus (Albrecht et al., 2002). In this study, we used the following formula for describing the CRF:

\[ R = \frac{R_{\text{max}} C_t^\alpha}{\sigma^\beta + C_t^\beta}. \]
Separate $\alpha$ and $\beta$ have been used for fitting the CRF of single cell data (Chen, Kasamatsu, Polat, & Norcia, 2001; Li & Creutzfeldt, 1984), VEP (Ross & Speed, 1991; Ross, Speed, & Morgan, 1993), and behavioral data (Xing & Heeger, 2000). For example, the amplitude of a response to a high contrast stimulus can be smaller than that of the response to a lower contrast stimulus, a phenomenon referred to as “oversaturation” (Li & Creutzfeldt, 1984; Regan, 1989; Sclar et al., 1990). Such data cannot be fitted by Equation 1, which is monotonic. Studies involving spatial interaction in the visual system of the monkey (Chen et al., 2001; Somers et al., 1998) have suggested that $\alpha$ and $\beta$ are related to the excitatory and inhibitory modulations, respectively.

When a neighbor stimulus is present, mutual inhibition of the responses to both target and neighbor stimuli is often observable. Two models, a simple multiplicative model and the normalization model, have been used to describe spatial interaction in the visual cortex. In cat visual cortex, Chen et al. (2001) studied how the response of a neuron is influenced by a neighbor stimulus, which does not excite the studied neuron when presented alone. For the cases where the neighbor stimulus inhibits the response of a neuron, the data suggest that the inhibitory influence of the neighbor reduces the response to the target by a constant multiplicative factor (Chen et al., 2001). However, other studies (Candy, Skoczynska, & Norcia, 2001; Heuer & Britten, 2002; Sceniak, Hawken, & Sharpley, 2001; Simoncelli & Schwartz, 1999; Simoncelli & Heeger, 1998; Tolhurst & Heeger, 1997; Xing & Heeger, 2001) support the normalization model (Carandini, Heeger, & Movshon, 1997; Heeger, 1993). The normalization model was originally proposed to describe the interaction among cortical channels, such as orientation and spatial frequency channels. Note that interactions among cortical channels differ from the spatial interaction: the visual stimuli used for studying interactions among cortical channels often overlap each other, while the stimuli used for studying spatial interaction should be separated. Nonetheless, the success of the normalization model indicates that features of the normalization model, such as mutual inhibition and response normalization, are important in describing spatial interaction. The normalization model assumes that the responses to multiple stimuli are pooled to generate a divisive inhibition. Although it appears in various forms in different papers, the normalization model is an extension of Equation 2 and can be expressed as (Heeger, 1993; Xing & Heeger, 2001):

$$R = \frac{R_{\text{max}} C^\alpha_t}{\sigma^\beta + C^\beta_t + kC^\beta_n},$$

where $C_n$ is the contrast of a neighbor stimulus, and $k$ is factor that determines the strength of the inhibitory effect.

The normalization model has been shown to be fairly consistent with a wide range of single cell recordings (Albrecht & Geisler, 1991; Sceniak et al., 2001; Simoncelli & Heeger, 1998) and psychophysical data (Chen, Foley, & Brainard, 2000; Foley, 1994). Notice that when $C_t$ is much larger than $C_n$, the effect of $C_n$ can be neglected. When $C_t$ is similar to $C_n$, $kC^\beta_n$ effectively is added to the $\sigma^\beta$ term, and thus the effective semi-saturation contrast is increased. This effect has been called a “contrast gain” change. In other words, spatial interaction changes the effective contrast of the target stimulus in the CRF, a result often found in electrophysiological and psychophysical studies (for a review, see Boynton, 2005; Kanwisher & Wojciulik, 2000; Reynolds & Chelazzi, 2004; Treue, 2001). In addition, the normalization model has been shown, with information theory, to allow the visual system to code nature images more efficiently (Schwartz & Simoncelli, 2001; Valerio & Navarro, 2003).

However, the spatial interaction results are often more complex than the normalization model predicts. For example, in the middle temporal region (MT), a neuron’s responses to a set of moving dots in a given target direction, when many dots were moving in another direction, is significantly higher than predicted by the normalization model (Simoncelli & Heeger, 1998; Snowden, Treue, Erickson, & Andersen, 1991). A similar effect also has been shown in psychophysical data (Ejima & Takahashi, 1985), where the inhibitory effect of the neighbor stimulus approaches an asymptotic level when the contrast of the neighbor stimulus is either much higher, or much lower, than the contrast of the target contrast. When two sinusoidal gratings of different orientations are superimposed, the CRF for the target stimulus, measured with the conventional VEP, clearly deviate from the predictions of a normalization model (Ross & Speed, 1991). As Carandini et al. (1997) pointed out, the normalization model does not appropriately describe the responses when the neighbor contrast is high.

One obstacle to a better understanding of spatial interaction is the difficulty of recording separate responses to the two simultaneously presented stimuli, the target and the neighbor stimuli. Interactions between two stimuli have been investigated with VEP techniques in which two stimuli were modulated with temporal sinusoidal function with different frequencies (Grose-Fifer, Zemon, & Gordon, 1994; Regan & Regan, 1988; Victor & Conte, 2000; Victor, Purpura, & Conte, 1998). These studies demonstrated that lateral interactions could be measured with the VEP. Here, we employ a multifocal visual evoked potential (mfVEP) paradigm, in which multiple visual stimuli are presented simultaneously and independently, and the response to each stimulus obtained. This method allows us to distinguish the visual responses to the target and neighbor stimuli. Another advantage of the mfVEP paradigm is that the
mfVEP response is largely generated in V1, unlike the conventional full field VEP, which has significant extrastriate components (Fortune & Hood, 2003; Slotnick, Klein, Carney, Sutter, & Dastmalchi, 1999; Zhang & Hood, 2004).

In summary, although the normalization model fits the data well in many cases, it is not an appropriate explication when the neighbor stimulus has a high contrast and the target stimulus has a low contrast. In this study, we systematically varied the contrasts of both the target and neighboring stimuli to provide a test of models of spatial interaction.

Method

The display

The visual stimulus was a pattern-reversing dartboard stimulus composed of one ring of 24 sectors, subtending 44.5° of visual angle. The sectors were interleaved with two contrasts (e.g., 4% and 16%). The dartboard pattern shown in Figure 1A provides an example of the display. We called the sectors at the 1st, 3rd,..., 23rd positions odd sectors and those at the 2nd, 4th,..., 24th positions even sectors.

![Stimulus](image)

Figure 1. (A) The one-ring, 24-sector display. The traces are the mfVEP responses, recorded for two minutes, attributed to each sector. The sector number is arranged in clockwise direction, starting at 3 o’clock. The sectors 1 and 2 and the correspondent responses are marked. Red and orange traces present the responses to the 16% stimuli, and blue and green ones are those to the 4% stimuli. (B) Both the mfVEP responses to 16% stimulus with neighbor contrast of 4% and the mfVEP responses to 4% stimulus with neighbor contrast of 16% were obtained by combining the data recorded in the two conditions.
sectors. The odd and even sectors served mutually as targets and neighbors to each other.

Multifocal VEP

An mfVEP stimulus is a dartboard display comprised of a number of sectors, each with a checkerboard pattern. Both the sectors and checks inside a sector vary in size with retinal eccentricity roughly according to the cortical magnification factor (Baseler, Sutter, Klein, & Carney, 1994). Each sector is an independent stimulus that reverses in contrast every 13.33 ms (for screen refresh rate of 75 Hz) according to an m-sequence, which is a pseudorandom sequence that allows for simultaneous modulation of many independent stimuli (Sutter & Tran, 1992). Due to the randomness of the m-sequence, the effective inter stimulus interval (ISI) covers a wide range and thus ensures effective stimulation of the visual system, even though each step of the m-sequence lasts only for 13.33 ms. The first slice of the 2nd order kernel, which is essentially the averaged response that is time-locked to the contrast reversing events, is obtained and referred to as the mfVEP response in this study. To ensure that the mfVEP responses are not contaminated with higher order kernels (Sutter, 2001), the average data for all subjects were inspected. The top right panel in Figure 2 presents the response to 0% stimulus when the neighbor contrast is 64%. Higher order kernel contributions, if they existed, should be seen clearly in this condition. The top right panel and the bottom left panel data were derived from the same recording, either the [0 64] condition or the [64 0] condition. The small response to 0%, when it is simultaneously obtained with the 64% response (shown in the bottom left panel), indicates that higher order kernel contamination, if present, is not significant.

The responses to the target and the neighbor stimuli

In Figure 1A, the traces surrounding the dartboard are the mfVEP responses, recorded in one session, averaged across the 4, 30-second runs. The amplitudes of the responses of the sectors are correlated to the contrast of them (4% and 16%). To obtain the response to the stimulus of one contrast when the neighbor contrast is fixed, the data from two conditions are combined (Figure 1B).

The experiment has a six-by-six design, where the contrasts used in both the target and surrounding stimuli were 0%, 4%, 8%, 16%, 32%, and 64%, a total of 36 conditions.

Recording

The mfVEP stimulus presentation and data analysis were performed with a custom c++ program (Zhang, 2003). This software allowed for the presentation of an mfVEP stimulus of desired shape and location. A 2048-step m-sequence, lasting 27 seconds, was used. Each subject was tested on two days; each day contained four sessions in which the 36 conditions were arranged in random order. The display is a CRT monitor with a refresh rate of 75 Hz. The electrodes that comprised the midline channel were placed at the inion (reference) and 4 cm above the inion (active) with a forehead electrode serving as the ground. Additional active electrodes were placed at 4 cm lateral (left or right) to the midline and 1 cm above the inion. The midline active electrode and the two lateral active electrodes were referenced to the inion electrode and created three recording channels. The positions of the active electrodes were based on anatomical considerations and chosen for optimizing mfVEP recording (Hood, Zhang, Hong, & Chen, 2002). The continuous VEP record was amplified with the high and low frequency cutoffs set at 3 and 100 Hz (Grass PreAmplifier P511J, Quincy, MA). The mfVEP responses were filtered offline with the high and low frequency cutoffs set at 3 to 35 Hz using a fast Fourier transformation (Hood et al., 2003).

The three subjects had normal vision corrected to 20/20. Informed consent was obtained from all subjects before
their participation. Procedures adhered to the tenets of the Declaration of Helsinki, and the protocol was approved by the committee of the Institutional Board of Research Associates of Columbia University.

The amplitude measure

Because the mfVEP amplitude varies across subjects due to differences in cortical convolution and skull conductance, we normalized the data to obtain relative amplitudes. For each subject and for each session, a template of the mfVEP responses was obtained by averaging the data for all conditions. The amplitude of mfVEP for one condition was calculated as the scalar product between the response and the template (Sutter & Tran, 1992) and then divided by the average amplitude of all 36 conditions. Therefore, the amplitude is a relative measure. This method was adopted because the amplitude of the response to a stimulus with low (e.g., 4%) contrast is very difficult to obtain and yet is valid because the waveform of mfVEP is not significantly influenced by the contrast of the visual stimulus (Hood et al., 2006).

The comparison between two competitive models

To determine the best of the competing spatial interaction models, we used the Akaike information criterion (AIC) (Akaike, 1973).

\[
AIC = n \ln \left( \frac{RSS}{n} \right) + 2k, \tag{4}
\]

where \( n \) is number of observations, RSS is residual sum of squares, and \( k \) is the number of parameters in the model.

AIC takes into consideration both the goodness of fit and the number of parameters that have to be estimated to achieve this particular degree of fit and therefore imposes a penalty for increasing the number of parameters.

Results

Figure 2 shows the average responses for all conditions. Each trace is average of either lower (red) and upper field (blue) mfVEP. Each row presents data for one target contrast, and each column presents for one neighbor contrast. In general, the amplitude of the mfVEP increases with the contrast of the target (from top to bottom) and decreases with the contrast of the neighbor (from left to right). This observation is detailed in Figure 3. Each row presents the data for one subject, and each point is obtained by combining data from two sessions with 4, 30-second runs per session. The two columns present the same data, the relative amplitude of the mfVEP response. The relative amplitude for one condition is calculated by dividing the amplitude by the average amplitudes for all conditions. On the left, the abscissa is the contrast of the target and each color represents the CRF when the neighbor contrast is a given value. Notice that as the target contrast is increased, the CRFs in Figure 3 (left column) tend to saturate, as sometimes found in mfVEP studies (Baseler & Sutter, 1997; Hood et al., 2006; Klistorner, Crewther, & Crewther, 1997). However, for higher neighboring contrasts, the CRFs appear to have a higher semi-saturation constant and a lower maximum response.

The dotted lines are the predictions of the normalization model (Equation 3), and the solid lines present a new model: a multiplicative spatial interaction model (Equation 5). The amplitude of the response \( R \) of this model is given by

\[
R = \frac{AC_t}{\sigma^\beta + C_t^\beta} \left( 1 + \frac{B}{1 + \left( \frac{AC_t}{C_t} \right)^\gamma} \right), \tag{5}
\]

where \( B \) is the factor describing the strength of the spatial interaction, \( \gamma \) is a power term that describes nonlinearity of the spatial interaction, and \( k \) is a factor that describes the effective contrast of the neighbor stimulus. Note that when \( C_t \) is zero, \( A(1 + B) \) is the \( R_{max} \) term in Equations 1, 2, and 3. Although it appears more complex, the spatial interaction term is a mathematical description of a sigmoid curve.

The smooth curves in Figure 3 (left column) show the fits of this model (solid) and the normalization model (dotted). Both models capture the effect on the magnitude of a response to a stimulus when affected by the contrast of the neighbor stimuli. However, when the contrast of neighbor stimulus is high (the data in red in the left column, 64% neighbor contrast conditions), the normalization model underestimates the data. The pattern of results for the 64% target condition (red in the left column of Figure 3) is similar to that observed in previous psychophysical (Ejima & Takahashi, 1985; Xing & Heeger, 2001), single cell (Snowden et al., 1991) and VEP studies using superimposed sinusoidal gratings (Ross & Speed, 1991).

The mechanisms of spatial interaction are better illustrated by the plots in the right column of Figure 3. Here, the abscissa is the contrast of the neighbor stimulus, while each color represents data for one target contrast. First, the effect of the neighboring contrast on the target
response sharply increases when the neighboring contrast is equal to, or larger than, the target contrast. Second, when the neighboring contrast is much higher than the target contrast (green, cyan, and blue data points in the right column of Figure 3), the response to the target stimulus is not further reduced, as if the spatial interaction effect reaches an asymptotic level. The multiplicative model (Equation 5) was constructed to describe these features of the results.

Table 1 lists the parameters of the multiplicative model for the fit to the data of the three subjects. The ΔAIC values indicate that the multiplicative model does better than the normalization model, even when taking into consideration the fact that the multiplicative model has two additional parameters.

Discussion

Our mfVEP data are consistent with the data reported by Ejima and Takahashi (1985) and Xing and Heeger (2001), in spite of the differences in methodology (mfVEP
vs. psychophysical) and visual stimuli (checkerboard vs. sinusoidal grating). Because the normalization model captures both the nonlinearity of the visual system and the reduction in response amplitude by the neighbor stimulus, it fits the spatial interaction data in previous works fairly well (Xing & Heeger, 2001). However, our data show that when the neighbor stimuli have a high contrast, the amplitude of the response to the target stimulus is consistently higher than that predicted by the normalization model. Since a similar phenomenon has been observed in many studies, our model may help to better interpret those data.

A close inspection of our data suggests that the discrepancy between data and the normalization model does not necessarily depend upon the high contrast of the neighbor stimuli. Instead, the discrepancy arises when the ratio between the neighboring contrast and target contrast is high. To describe our mfVEP data, we provide an alternative model where the spatial interaction is a multiplicative process that is determined by the contrast ratio between the two stimuli. Our model suggests a different mechanism of spatial interaction than that depicted in the normalization model. First, our model describes the spatial interaction as a multiplicative process. The spatial interaction term and the physical contrast of the target stimulus are separate terms that are multiplied together to determine the amplitude of the target response. Second, the spatial interaction mechanism is nonlinear. The $\gamma$ term in Equation 5 is larger than 1. Therefore, when $C_j/C_t$ deviates slightly from 1.0 ($C_j/C_t)^\gamma$, and the spatial interaction term will change dramatically. This reflects the mutual inhibition between target and neighbor, where the stimulus with the slightly larger contrast exerts a much stronger influence than predicted by the difference in contrasts of the two stimuli. Consequently, the difference between target contrast and the neighbor stimuli is amplified. Third, the multiplicative model emphasizes the saturation of the spatial interaction when two stimuli have very different contrasts. Therefore, a weak target stimulus among strong neighbor stimuli can remain visible because the spatial inhibition from the neighbor response is limited. In contrast, the normalization model describes the divisive inhibition as $C_n^\beta$, where $\beta$ is larger than 1. Therefore, the normalization model predicts that the target response will approach zero when $C_n$ is large.

The normalization model is appealing for its mathematical simplicity and because it accounts for both the nonlinearity of the contrast response function and the mutual inhibition among multiple responses. However, previous studies (Carandini et al., 1997; Ejima & Takahashi, 1985; Ross & Speed, 1991; Snowden et al., 1991) showed that the normalization model underestimates the target response when the neighbor contrast is high and the target contrast is low. Our data confirmed this observation. In the normalization model, both the response to the target and that to the neighbor stimulus are pooled together to form the mutual inhibition mechanism, therefore suggesting that a neuron and its neighbor neurons share a common negative feedback mechanism. The better fit of our model to the data suggests that such a tight coupling between the neighboring neurons may not be appropriate for describing spatial interaction, perhaps neurons are separated at a greater distance. A contrast gain effect has been reported in other studies such as those examining visual spatial attention (Reynolds, Pasternak, & Desimone, 2000). Since a spatial attention experiment often involves distracter stimuli, the multiplicative model may also help us to better understand the mechanism of spatial attention.

In conclusion, although spatial interaction data does not seem to support a simple (multiplicative) model, our analysis suggests that spatial interaction can take place though a multiplicative process, when the nonlinearity and the mutual inhibition involved in spatial interaction are taken into consideration. This multiplicative model might provide a new platform for analyzing the issues related to the spatial interaction.

### Acknowledgments

This work was supported by the National Eye Institute of the National Institute of Health Grant R01-EY-02115 (DCH) and Dana Foundation.

Commercial relationships: none.

Corresponding author: Xian Zhang.

Email: xzhang63@gmail.com.

Address: 514 West, 114th St., Apt 3B, New York, NY 10025, USA.
References


