The what and where in visual masking

Haluk Ogmen, Bruno G. Breitmeyer, Reginald Melvin

Abstract

A metacontrast mask suppresses the visibility of, without influencing the reaction time (RT) to, the target. We investigated whether this dissociation results from a sensori-motor pathway immune to masking effects or from the characteristics of stimulus timing in mutually inhibitory sustained and transient channels. For target visibility, para- and metacontrast yielded the usual U-shaped functions. Peak paracontrast occurred at stimulus onset asynchronies (SOAs) of 150 to 100 ms. RTs were relatively low for metacontrast and did not show a systematic change as a function of SOA. The RT contribution from contour-masking was greatest at an SOA of 150 ms (paracontrast) and declined to near zero in the metacontrast regime. The dissociation between visibility and RT seen in metacontrast did not occur in paracontrast, rejecting the theory that RTs are elicited by a single sensori-motor pathway immune to masking. The dependence of the dissociation on stimulus timing can be explained by RECOD, a dual-pathway model wherein fast and slow activities interact.

Keywords: Paracontrast; Metacontrast; Visual masking; Dissociation

1. Introduction

The visibility of a target stimulus can be strongly reduced when it is followed in time by a spatially non-overlapping mask stimulus, a phenomenon known as metacontrast (Fry, 1934; Stigler, 1910, 1926; reviews: Bachmann, 1994; Breitmeyer, 1984; Breitmeyer & Ogmen, 2000). In general, the interactions between the target- and mask-generated activities are complex and the criterion content, i.e. the stimulus dimension(s) on which the observer bases his/her judgments, has profound effects on the nature of masking effects (reviews: Kahneman, 1968; Breitmeyer, 1984, pp. 103–105). For example, for target and mask stimuli of equal energy, a U-shaped (“Type B”) metacontrast function is obtained when the observers make judgments related to the target’s surface properties (e.g., perceived brightness, contrast), contour properties (e.g., contour completeness, contour shape), or figural identity (e.g., letter recognition). U-shaped functions can be obtained by a variety of response tasks such as matching, magnitude estimation, and choice reaction-times (RTs) as long as the criterion content required for the task falls into one of the aforementioned categories (for references, see the aforementioned reviews). However, when the observer’s task is changed to report the presence or the spatial location of the target, instead of its visibility, the metacontrast mask has no effect on the observer’s performance, as measured by simple/choice RTs or by response accuracy (e.g., Fehrer & Raab, 1962; Schiller & Smith, 1966). A “coarse localization” of the stimulus

Note that, as one might expect, in addition to task parameters and criterion content, stimulus parameters such as size, energy, eccentricity, contour proximity, also influence the nature of masking interactions. For example, when the mask energy is higher than the target energy one obtains a monotonic (“Type A”) masking function. The absence of masking effects in Schiller and Smith’s study held when the target and mask stimuli were of equal energy. When the mask was of higher energy than the target, a Type A masking function was obtained (Schiller & Smith, 1966).
in time and/or in space is sufficient to accomplish this task. We will therefore make a distinction between task parameters and criterion content that require a coarse localization of the target (hereafter referred as “target localization”) and those that require surface, contour, and figural identification of the target (hereafter referred as “target visibility”). 2 The markedly different findings for the visibility versus localization of the target suggest that the processes that underlie the visibility of a stimulus can be dissociated from those that underlie the spatial localization of the same stimulus.

Several studies provided evidence for the existence of two distinct neural pathways, a “what” pathway related to stimulus visibility/identity and a “where” pathway related to the spatial localization of the stimulus (Desimone & Ungerleider, 1989; Ungerleider, 1985; Ungerleider & Mishkin, 1982). Accordingly, one way to explain the dissociation between stimulus visibility and localization in visual masking is to postulate that the mask stimulus interferes with the target stimulus in the “what” pathway but not in the “where” pathway. 3 This explanation would be consistent with the findings showing that figural aspects of stimuli such as size can be distorted at the perceptual level but not at the motor level (Milner & Goodale, 1995). On the other hand, it is also possible that the pathways processing the figural and the spatial localization information interact and that the dissociation observed in metacontrast is not a general property of neural processing. Indeed, our analysis of a dual-channel model of masking (explained in detail in Section 2) suggests that no dissociation should be observed when the temporal order of the stimuli is reversed (the mask is presented before the target, i.e. paracontrast). The purpose of this study was to determine whether the dissociation between visibility and spatial localization in masking holds irrespective of stimulus timing and test whether the dissociation results from a sensori-motor pathway immune to masking effects or from the characteristics of stimulus timing in mutually inhibitory sustained and transient channels.

2. Theoretical background

2.1. The general architecture of the model

In this section, we discuss the predictions of the retino-cortical dynamics (RECOD) (Ogmen, 1993) model for stimulus visibility and localization. Fig. 1 illustrates the general structure of the model.

Two major populations of ganglion cells, one with fast-phasic (transient) responses and a second with slower-tonic (sustained) responses had been identified in the primate retina (e.g., Croner & Kaplan, 1995; De Monasterio, 1978; Gouras, 1969). The two ellipses at the bottom of Fig. 1 represent these two populations of retinal ganglion cells in our model. Typical responses of these neurons to a pulse input are depicted in the figure. These two populations of retinal ganglion cells project to distinct layers of the lateral geniculate nucleus (LGN) and form two parallel afferent pathways (the magnocellular and the parvocellular) as shown in the figure. At the systems level, the properties of transient and sustained channels in humans (Breitmeyer, 1975, 1984; Breitmeyer, Levi, & Harwerth, 1981; Kulikowski & Tolhurst, 1973; Legge, 1978) and monkeys (Harwerth, Boltz, & Smith, 1980) parallel the properties of magnocellular and parvocellular pathways, respectively (Kremers, 1999) and we consider these pathways as neural correlates for the transient and sustained afferents in our model. Magnocellular and parvocellular projections to the cortex provide selective inputs to different visual areas sub-serving various functions such as the computation of motion, form, and brightness. However, at the cortical level these two pathways interact (Van Essen, Anderson, & Felleman, 1992), but the loci and degree of their interactions are not fully established (e.g., Martin, 1992; Sincich & Horton, 2002). Neuro-anatomical data indicate that the magnocellular afferents show that the processes that underlie the visibility of a stimulus can be dissociated from those that underlie the spatial localization of the same stimulus.

Several studies provided evidence for the existence of two distinct neural pathways, a “what” pathway related to stimulus visibility/identity and a “where” pathway related to the spatial localization of the stimulus (Desimone & Ungerleider, 1989; Ungerleider, 1985; Ungerleider & Mishkin, 1982). Accordingly, one way to explain the dissociation between stimulus visibility and localization in visual masking is to postulate that the mask stimulus interferes with the target stimulus in the “what” pathway but not in the “where” pathway. 3 This explanation would be consistent with the findings showing that figural aspects of stimuli such as size can be distorted at the perceptual level but not at the motor level (Milner & Goodale, 1995). On the other hand, it is also possible that the pathways processing the figural and the spatial localization information interact and that the dissociation observed in metacontrast is not a general property of neural processing. Indeed, our analysis of a dual-channel model of masking (explained in detail in Section 2) suggests that no dissociation should be observed when the temporal order of the stimuli is reversed (the mask is presented before the target, i.e. paracontrast). The purpose of this study was to determine whether the dissociation between visibility and spatial localization in masking holds irrespective of stimulus timing and test whether the dissociation results from a sensori-motor pathway immune to masking effects or from the characteristics of stimulus timing in mutually inhibitory sustained and transient channels.

2.1. The general architecture of the model

In this section, we discuss the predictions of the retino-cortical dynamics (RECOD) (Ogmen, 1993) model for stimulus visibility and localization. Fig. 1 illustrates the general structure of the model.

Two major populations of ganglion cells, one with fast-phasic (transient) responses and a second with slower-tonic (sustained) responses had been identified in the primate retina (e.g., Croner & Kaplan, 1995; De Monasterio, 1978; Gouras, 1969). The two ellipses at the bottom of Fig. 1 represent these two populations of retinal ganglion cells in our model. Typical responses of these neurons to a pulse input are depicted in the figure. These two populations of retinal ganglion cells project to distinct layers of the lateral geniculate nucleus (LGN) and form two parallel afferent pathways (the magnocellular and the parvocellular) as shown in the figure. At the systems level, the properties of transient and sustained channels in humans (Breitmeyer, 1975, 1984; Breitmeyer, Levi, & Harwerth, 1981; Kulikowski & Tolhurst, 1973; Legge, 1978) and monkeys (Harwerth, Boltz, & Smith, 1980) parallel the properties of magnocellular and parvocellular pathways, respectively (Kremers, 1999) and we consider these pathways as neural correlates for the transient and sustained afferents in our model. Magnocellular and parvocellular projections to the cortex provide selective inputs to different visual areas sub-serving various functions such as the computation of motion, form, and brightness. However, at the cortical level these two pathways interact (Van Essen, Anderson, & Felleman, 1992), but the loci and degree of their interactions are not fully established (e.g., Martin, 1992; Sincich & Horton, 2002). Neuro-anatomical data indicate that the magnocellular afferents...
provide the dominant (as opposed to exclusive) inputs to the dorsal ("where") pathway whereas parvocellular afferents provide the dominant inputs to the ventral ("what") pathway (Yabuta & Callaway, 1998). The model uses a lumped representation for the cortical targets of magnocellular and parvocellular pathways. The main cortical targets of the magnocellular pathway represent the areas in the dorsal pathway. The main cortical targets of the parvocellular pathway represent the areas in the ventral pathway (see the upper ellipses in Fig. 1). The model postulates that reciprocal inhibition exists between parvo and magno driven cortical cells as shown by the arrows between the upper ellipses in the figure. We will refer to this reciprocal inhibition as inter-channel inhibition to distinguish it from inhibitory interactions within each channel, which in turn will be called intra-channel inhibition (Breitmeyer, 1984). The lumped representation for the areas involved in the computation of dynamic form and brightness contains recurrent connections to represent the extensive feedback observed between cortical areas as well as the feedback from cortex back to LGN (rev. Sherman & Guillery, 1996). The real-time dynamics of the model unfolds in three phases: (i) a “feed-forward dominant” phase where the afferent signals travel to higher cortical areas, (ii) a “feedback dominant” phase where feedback, or re-entrant, signals contribute to information processing, and (iii) a “reset phase” which allows a transition from feedback back to feed-forward dominant phase when inputs change. An illustration of these phases can be seen in Fig. 3 of Purushothaman, Ogmen, Chen, & Bedell (1998).

2.2. Schematic explanation of model predictions

A detailed description of the model can be found in Appendix A. In this section, we will present a simplified schematic explanation of the model predictions in order to provide an intuitive basis for the quantitative results presented in the manuscript. The top panel of Fig. 2 provides a schematic explanation for the dissociation between visibility and localization in metacontrast. The target stimulus is presented first and generates a fast transient and a slower sustained activity in the afferent transient and sustained pathways, respectively. The model postulates that the visibility of the target is correlated with the activities in post-retinal areas receiving their main input from the sustained (parvocellular) pathway. Both sustained and transient activities carry information about the spatial location of the target. However, because of the shorter latency of the transient signals (Maunsell & Gibson, 1992; Petersen, Miezin, & Allman, 1988; Schmolesky et al., 1998), the model postulates that the transient activities will play a major role in target localization when the observer is asked to respond as fast as possible. This provides a theoretical rationale as to why RT is used as the dependent variable in our localization experiments.

In metacontrast, the mask is presented second and generates similar activities with a delay equal to the stimulus onset asynchrony (SOA). From the nature of the temporal overlap between the activities, one can see that both the intra- and inter-channel inhibition will cause a suppression of activity in the sustained channel for the target. As a result, the visibility of the target is predicted to decrease. However, the transient activity generated by the target, and consequently the ability of the observer to report the presence or the location of the target, is predicted to remain intact.

The bottom panel of Fig. 2 depicts the prediction of the model for paracontrast. In this case, both the transient and sustained activities generated by the target are inhibited. As a result, the model predicts that RTs should increase and the visibility of the target should decrease.

4 In general, excitatory connections between these areas are required to model cross-attribute interactions such as motion-defined form. However, cross-attribute interactions are not critical in the analysis presented herein and thus, for simplicity, we did not elaborate the model to include inter-channel excitatory connections.
2.3. Model simulations

As discussed in Appendix A, all model equations and parameters, with a few exceptions, were taken from our previous studies (Ogmen, 1993; Purushothaman et al., 1998; Purushothaman, Lacassagne, Bedell, & Ogmen, 2002). The network was designed to represent spatially a one-dimensional field of 3.2 deg around the fovea. A spatially faithful representation of the stimuli used in this study would require a much larger network. Therefore, in our simulations we used “representative stimuli” consisting of one-dimensional versions of the actual stimuli designed to fit into the 3.2 deg field around the fovea. The model predicted reaction-time was determined as a weighted average of the latencies for the transient and sustained responses. In the simulations, under normal conditions the magnitude of the transient responses was much higher than that of the sustained responses. This difference biased the weighted-average towards the faster response, as required by the experimental task. Because the experimental task required the determination of a choice RT resulting from a discrimination of the activity generated by the mask-only versus by the target-and-mask in temporal succession, the transient and sustained activities used in determining the choice RT were those derived by computing the differential activities between mask-only versus target-and-mask cases. The model predicted perceived-brightness was computed as the time-integrated activity of post-retinal sustained cells responding to the target.

Fig. 3 shows the results of the simulations. Type B, or U-shaped masking functions are obtained for both paracontrast and metacontrast. For target localization, RTs are relatively constant for metacontrast but increase for paracontrast.

A comparison of these results with the corresponding data will be presented in the following sections.

3. Stimulus visibility in paracontrast and metacontrast

While stimulus visibility in metacontrast had been studied extensively, there have been very few studies of paracontrast. Therefore, we first measured target visibility for both metacontrast and paracontrast.

3.1. Methods

Apparatus and stimuli. The stimuli were displayed at 160 Hz frame rate on a NANA0 Flex-Scan F2-21 color monitor. Stimulus presentation and response recording were controlled by a visual stimulus generator (VSG2/3) card manufactured by Cambridge Research Systems. Fig. 4(A) illustrates the stimulus configuration. A fixation cross was presented at the center of the screen. The target was a disk of 1 deg diameter. The center of the target was displaced horizontally 1.7 deg to the right of the fixation square. The mask consisted of a ring of 1.7 deg outer diameter. The spacing between the target disk and the mask ring was 6 min. A match disk of 1 deg diameter was placed 1.7 deg horizontally to the left of the fixation target. No mask stimulus surrounded the match disk. The background luminance was 8.5 cd/m². The target and mask stimuli had the same luminance of 27.5 cd/m². The target and mask stimuli had the same luminance of 27.5 cd/m². The target, mask, and the match stimuli had all the same duration of 12.5 ms. The target and the match were always presented simultaneously. The mask was presented with SOA values which were ordered randomly during the experiment. To establish a baseline value for target visibility, we also ran a condition where the mask was not presented (T-only).

Observers. Two of the authors (RM was naive to the purpose of the experiment at the time of data collection) participated in the experiment.

Procedure. The task of the observer was to indicate through a joystick which of the two, the target or the match, appeared brighter. The luminance of the match...
varied according to a staircase procedure to estimate the point of subjective equality (PSE).

3.2. Results and discussion

Fig. 5 shows the PSEs between the perceived luminance of the match and the target stimuli as a function of SOA for the two observers. As expected, both para and metacontrast functions show U-shaped functions with a smaller magnitude for para than metacontrast. Peak metacontrast occurs at SOA = 50 ms. Peak paracontrast occurs in the ~150 to ~100 ms range, which is similar to the findings obtained using chromatic and monocular stimuli (Cavonius & Reeves, 1983) but substantially larger than reports showing peak paracontrast in the ~30 to ~20 ms range using dichoptic stimuli (Kolers & Rosner, 1960). It is known that peak metacontrast magnitude and timing vary with stimulus parameters and the observer’s task (Breitmeyer, 1984, pp. 105–120). It is also possible that paracontrast timing can vary with stimulus parameters and the observers’ task. As mentioned earlier, there have been very few studies of paracontrast and further studies are needed to answer this question.

Fig. 6 shows the data averaged across the two observers along with the model prediction. Overall, the model predictions match well the data in terms of the location of the dips in the masking function. In the model, paracontrast masking at relatively large SOA magnitudes arises from a slow intra-channel cortical inhibition which was not included in the earlier versions of the dual-channel masking models. In our simulations, the relative delay of this intra-channel inhibition was 144 ms. The estimated delays for antagonistic interactions (center-surround) in the early visual pathways are at least an order of magnitude less than this value (e.g., Benardete & Kaplan, 1997). Both the type and timing of post-retinal interactions are complex, but if our assumption that this inhibition is carried out through intra-channel mechanisms is correct, consideration of response timings in the visual system (Schmolesky et al., 1998) suggests the involvement of anatomically efferent signals. In our simulations, we used a functionally feed-forward signal to implement this inhibition (see Fig. 7); however, a functionally feedback signal is also possible.

Fig. 5. The averaged perceived brightness of the target as a function of SOA for two observers. The error bars correspond to ±1 SEM. The horizontal lines represent the perceived brightness of the target in the absence of the mask stimulus.

Fig. 6. Data from Fig. 5 averaged across the two observers and normalized with respect to the target-only condition ±1 SEM superimposed on the model predictions from Fig. 3.
Additional experiments are needed to test these possibilities. Quantitatively, the model under-estimates the magnitude of metacontrast and the span of masking at large SOA magnitudes. For computational simplicity, the simulations were based on a simplified (one-dimensional) version of the stimuli used in the experiment. It is possible that the quantitative discrepancies between the model and the data are due to the failure in our model and simulations to represent adequately all stimulus parameters such as eccentricity, stimulus size, energy, and background level. These parameters are known to influence both the magnitude and the morphology of the masking functions (rev. Breitmeyer, 1984). Qualitatively, the model shows changes in stimulus visibility for both paracontrast and metacontrast and thus provide the necessary conditions to study the dissociation phenomenon. Our goal was to fit the model to the visibility data with a reasonable match so that quantitative predictions that provide a reasonable approximation for RT data could be obtained.

4. Spatial localization in paracontrast and metacontrast

In this section we introduce experiments that measure the performance of observers in target spatial localization in order to test whether the dissociation between target visibility and spatial localization holds in paracontrast. The parameters for the spatial localization experiment were slightly different than those used in the target visibility experiment (Section 3). First, we ran a control experiment to check that these parametric differences did not cause any dramatic change in the masking function. Second, we ran a pilot experiment whose results provided the rationale for the design of the main experiment.

4.1. Control and pilot experiments

4.1.1. Methods

Apparatus and Stimuli. The stimuli were displayed at a 75-Hz frame rate on a Sony Trinitron, 1024 × 768, color monitor. Stimulus presentation and response recording were controlled by a Macintosh II-ici computer. RTs were recorded by a National Instruments NB-MIO-16 input/output board with an onboard AM9513 timer running at 1 kHz. At a viewing distance of 57 cm, the display screen was ≈35 deg × 27 deg. As shown in Fig. 4B, a fixation cross appeared at the center of the screen, and the target and the mask stimuli were centered symmetrically 2 deg to the left and right of the vertical meridian and 1.6 deg above fixation. The background luminance was 100 cd/m². Stimuli were darker than the background. The target consisted of a dark disk of −33% contrast (50 cd/m²) and 0.86 deg in diameter. The mask ring could assume contrasts of −33% (50 cd/m²) or −100% (0 cd/m²). The outer diameter of the mask ring was 1.66 deg, and the contour separation between the disk and the mask ring was 2 min. Target and mask stimuli were each presented in separate frames, each lasting 13.33 ms. SOAs ranged from −293 to 224 ms in multiples of 13.33 ms.

Observers. Control and pilot experiments were conducted with one of the authors serving as the observer.

Procedure. Because there were some differences in the stimulus parameters with respect to the previous experiment, we conducted a control experiment to establish that changes in target visibility occurred for both paracontrast and metacontrast with the stimulus parameters used in this experiment. We used the method of magnitude estimation and the task of the observer was to rate the perceived brightness of the target on a scale ranging from 0 to 5, corresponding to the lowest and highest levels of visibility, respectively. Twenty ratings were averaged to provide an estimate of the perceived brightness of the target in the $M/T = 1$ condition. In a separate (pilot) experiment, we adopted a variation of the procedure used by Schiller & Smith (1966) to study RTs in metacontrast. On any trial, the target was randomly presented either to the left or else to the right of the vertical meridian and the observer had to press either the left or else the right arrow key of a keyboard as quickly and as accurately as possible to indicate where he detected the target. For metacontrast, nine SOAs ranging in 26.67 ms steps from 13.33 to 223.67 ms and for paracontrast, 10 SOAs ranging from −13.33 to −293.33 ms were used. For both the paracontrast and the metacontrast conditions the masks had either the
same contrast as the target or else three times the contrast of the target; i.e. mask-to-target contrast ratios were 1.0 or else 3.0. Metacontrast and paracontrast choice RTs were each run in separate experimental sessions, each consisting of four blocks of trials. The order of the four possible conditions consisting of crossing two mask-to-target contrast ratios (1.0 and 3.0) were randomized across blocks.

For metacontrast each of the blocks consisted of 180 trials with 20 trials devoted to each of the nine SOAs. Similarly, for paracontrast each of the blocks consisted of 200 trials with 20 trials devoted to each of the 10 SOAs. The orders of SOAs and target locations were randomized within a block. At each SOA, in half of the 20 trials the target was presented to the left of the vertical meridian; in the remaining half, to the right. Each trial began with a brief warning tone followed 600 ms later by the target-mask (for metacontrast) or the mask-target (for paracontrast) sequence. A millisecond clock was started with the presentation of the target and was terminated when the observer made his response. The subsequent trial began 1000 ms after the observer’s response. For each trial the computer stored the RT and whether the response was correct or incorrect.

4.1.2. Results and discussion

Fig. 8 plots the normalized ratings for target visibility (rating/5) along with normalized PSEs [(PSE in the masking condition)/(PSE in the target only condition)] from the previous experiment for observer BB. The two curves match well with the following exceptions: 6 For the conditions of the present experiment, for metacontrast the masking effect decays somewhat more slowly; paracontrast masking is stronger and seems to exhibit an additional component at SOA = −13.3 ms. The timing of this component is closer to the timing of paracontrast reported in the previous studies (Kolers & Rosner, 1960). 7 Overall, the results of the control experiment show that changes in stimulus parameters did not significantly modify the reduction observed in target visibility for both paracontrast and metacontrast.

Fig. 9 plots RTs as a function of SOA. The average response accuracy was 98%. As expected, RTs in metacontrast did not show a significant systematic dependence on SOA. In paracontrast, RTs were approximately constant for SOAs smaller than −100 ms at levels comparable to metacontrast. However, RTs increased significantly as SOA approached 0 ms. This pattern deviates from our model prediction. One factor which is not considered in our model is the “interference effect” that occurs when stimuli are presented in rapid succession. We hypothesized that besides contour-specific suppression of the target by the mask, the mask also produces an interference effect and, assuming additivity, the total reaction time RT_total can be expressed as:

\[ RT_{total} = RT_{contour-mask} + RT_{interference} + C, \]

where \( C \) is the baseline reaction time, \( RT_{contour-mask} \) and \( RT_{interference} \) represent the contributions of contour-mask and interference effects to the reaction time, respectively. \( RT_{contour-mask} \) can be estimated by designing a control experiment where a “pseudo-mask” produces an
interference effect with minimal contour masking. The reaction time obtained in the control experiment, $\text{RT}_{\text{control}}$, can be written as:

$$\text{RT}_{\text{control}} = \text{RT}_{\text{interference}} + C.$$ 

It can seen that the difference

$$\Delta \text{RT} = \text{RT}_{\text{total}} - \text{RT}_{\text{control}}$$

provides an estimate for $\text{RT}_{\text{contour-mask}}$.

4.2. The main experiment

4.2.1. Methods

All methods were identical to those described in Section 4.1 with the following changes to control for “interference” effects on RT:

Stimuli. A new “pseudo-mask” condition was introduced. The ring mask stimuli were replaced by four squares as shown in Fig. 4C. Each of the four squares comprising the pseudo-mask was 0.6 deg × 0.6 deg, and the horizontal and vertical distances from the center of the target to the center of each square were 1.7 deg. These parameters were chosen, based on our observations, so that the pseudo-mask, though having the same total area as the real mask, did not cause any significant change in the perceived brightness of the target.

Observers. The authors and four naive subjects obtained from the University of Houston undergraduate population served as observers. Of the naive observers, one was female. All subjects had normal or corrected-to-normal vision.

Procedure. Metacontrast and paracontrast choice RTs were each run in separate experimental sessions, each consisting of four blocks of trials. The order of the four possible conditions consisting of crossing two mask-to-target contrast ratios (1.0 and 3.0) and two mask types (real and pseudo) were randomized across blocks and counterbalanced across observers. For metacontrast, each of the blocks consisted of 180 trials with 20 trials devoted to each of the nine SOAs. Similarly, for paracontrast each of the blocks consisted of 200 trials with 20 trials devoted to each of the 10 SOAs. The orders of SOAs and target locations were randomized within a block. Observers ran the two—metacontrast and paracontrast—sessions on separate days. Each session lasted about 1 h. Order of session was counterbalanced across observers. At the beginning or end (randomized across observers) of the metacontrast sessions 20 choice RTs in which only the unmasked target was presented were measured. These choice RTs provided baseline data. Additionally, to compare single-subject to group performance, two of the authors (BB and HO), unlike the rest of the observers who each ran only two sessions, ran a total of eight sessions, repeating each of the metacontrast and paracontrast sessions four times in counterbalanced order. Across all possible conditions, observers BB and HO thus generated 80 responses per SOA instead of the 20 responses generated by the rest of the observers.

4.2.2. Results and discussion

The RTs and $\Delta$RTs averaged across the observers are plotted as a function of SOA in Figs. 10 and 11, respectively. The accuracy of responses was higher than 97% for all observers. As can be seen from Fig. 10, RTs asymptote to a somewhat higher level than the baseline RT, shown by the dashed line, for large magnitudes of SOA. This indicates that the mask adds a component to the baseline RT even when the target and the mask are well separated in time. For metacontrast, RTs are relatively flat showing some increase near SOA = 0, in agreement with previous findings (Schiller & Smith,
The increase near SOA = 0 is likely to be due to the interference effect. For paracontrast, RTs increase significantly as SOA approaches zero. Statistical analysis of the RT data showed that in paracontrast the effects of SOA, mask-type, and SOA × mask-type interaction were significant ($F_{9,45} = 45.9, p < 0.001; F_{1,5} = 39.0, p < 0.001; F_{9,45} = 3.4, p < 0.003$). SOA × mask-type interaction showed a significant quadratic trend ($F_{1,5} = 18.1, p < 0.008$). For metacontrast, only the effect of SOA was significant ($F_{8,40} = 7.9, p < 0.001$). RTs for the “real-mask” and “pseudo-mask” conditions show similar qualitative trends but with quantitative differences. Fig. 11 plots ARTs, i.e. the estimate of the “contour-mask” RTs. For metacontrast, ARTs fluctuate around averages of -5.5 and 1.7 ms for $M/T$ ratios of 1 and 3, respectively. However, for paracontrast ARTs depend strongly on SOA, peaking at SOA = -150 ms. The peak ART values are 28.7 and 51.1 ms for $M/T$ ratios of 1 and 3, respectively.

Overall, these results show that the dissociation between stimulus visibility and spatial localization does not hold in paracontrast. Fig. 11 also shows the model prediction superimposed on the data. Both the data and the model show an inverse-U shaped function for paracontrast and a relatively constant function for metacontrast. For paracontrast, on a closer examination, the model and the data (in particular for $M/T = 3$) suggest a finer structure that can be described as an “inverse-W function”, although the peaks and dips of this function in the model and the data are shifted with respect to each other.

Recently a related choice RT study employing metacontrast (Lachter & Durgin, 1999) showed that the ability to discriminate the location of a target-mask sequence from the location of a simultaneously presented two-mask sequence depended on task specification. When observers were allowed to make slow responses, they were least accurate at SOAs of 40–50 ms and progressively more accurate as progressively shorter or larger SOAs. In other words, a Type B non-monotonic U-shaped masking function was obtained. However, when required to make speeded responses, the accuracy was lowest at a 0-ms SOA and increased with SOA. Here a Type A monotonic function was found. We believe that these results can be explained by shifts in criterion content or effective visual information (Lachter & Durgin, 1999) used with the two task requirements. When allowed to make slow responses, the observers’ criterion could be based on the output of the slower sustained channels. Due to transient-on-sustained (interchannel) inhibition, these target-activated channels are suppressed in a U-shaped manner as a function of SOA by the fast mask-activated transient channels; hence one would expect a U-shaped masking function. However, when required to speed their responses, observers relied on the output of the faster target-activated and mask-activated transient channels. Here intrachannel interactions would prevail; and since the mask had a higher energy than the target, a Type A function would be expected, similar to the Type A function reported by Schiller & Smith (1966) in their target-localization task and also evident in our Fig. 10 at the shorter metacontrast SOAs of 0–90 ms.

5. Conclusions

We replicated prior findings showing that target-detection RTs (and ARTs) remain relatively constant in metacontrast. We have also shown that in paracontrast ARTs follow an inverse-U (or inverse-W) shaped function of SOA. We suggest that the difference between para and metacontrast originates from whether the target-generated transient activity is inhibited or not: According to our model, in metacontrast, target-generated transient activity remains intact and generates fast motor responses. In paracontrast, target-generated transient activity is inhibited, causing an increase in ARTs.

Overall, these findings suggest that the dissociation between stimulus visibility and spatial localization in visual masking does not originate from a sensori-motor pathway immune to masking effects. Rather, the dissociation depends on the timing of interactions between parallel pathways.

We have also found that strong paracontrast suppression occurs at SOAs substantially larger than those reported in some previous studies. Further studies are needed to clarify the origins of this difference.

Acknowledgements

This work was supported by NSF grant BCS-0114533 and NIH grant R01-MH49892. We thank the reviewers for helpful suggestions.

Appendix A. Mathematical description of the RECOD model

A.1. Introduction—fundamental equations of the model and their neurophysiological bases

The first type of equation used in the model has the form of a generic Hodgkin–Huxley equation

$$\frac{dV_m}{dt} = -(E_p + V_m)g_p + (E_d - V_m)g_d - (E_h + V_m)g_h,$$

(A.1)

where $V_m$ represents the membrane potential, $g_p$, $g_d$, $g_h$ are the conductances for passive, depolarizing, and hyperpolarizing channels, respectively, with $E_p$, $E_d$, $E_h$.
This system can be written as:

when the active state $X$ decays very fast, the dynamics of

turns dissociates back into $S$ and $Z$. It can be shown that

carries the signal to the next processing stage. This ac-

interacts with a transducing agent, $Z$, (e.g., a neuro-

where a biochemical agent, $S$, activated by the input,

have extensively in neural modeling to characterize

the dynamics of membrane patches, single cells, as well

as networks of cells (rev. Grossberg, 1988; Koch &

Segev, 1989). For simplicity, we will assume

as networks of cells (rev. Grossberg, 1988; Koch &

the dynamics of membrane patches, single cells, as well

been used extensively in neural modeling to characterize

representing their Nernst potentials. This equation has

been used in a variety of neural models, in particular to represent temporal adaptation,

or gain control property, occurring for example through

synaptic depression (e.g., Grossberg, 1972; Carpenter &

Grossberg, 1981; Ogmen & Gagne, 1990; Gaudiano,

1992; Ogmen, 1993; Abbott, Varela, Sen, & Nelson,

1997).

A.2. The retinal network

The retinal network is designed to capture the basic

spatio-temporal properties of the retinal output without

necessarily incorporating all details of the retinal cir-

uity. To the extent possible, parameters of the model

reflect the physiologically measured parameters of the

primate retina.

A.2.1. Retinal cells with sustained activities (parvocellular pathway)

All the equations and the parameters are identical to

those used in (Purushothaman et al., 2002). The activities

of sustained retinal cells are described in three

functional stages:

Stage I: Temporal adaptation (gain control). We use

Eq. (A.4) to achieve temporal adaptation (gain control)

where $z_i$ represents the concentration of a transducing

agent at the $i$th spatial location. $J$ is a baseline input

generating a dark current and $I_i$ is the external input

(luminance value) at the $i$th spatial position. This tem-

poral adaptation, or gain control stage causes the

neural activity to decay to a plateau level after an initial

peak response to a sustained input, as observed in sus-

tained retinal ganglion cell responses. The parameter $\tau$

adjusts the time-constant of the decaying response.

Stage 2: Spatial center-surround organization. Signals

from the first stage are convolved by the kernels $G^e_i$ and

$G^i_i$ which represent the excitatory-center and the inhibi-

tory-surround of the receptive field. The kernels are

Gaussian functions of the form $G^e_i = \text{Ampse}_i e^{-(x^2/(2\sigma^2))}$

and the parameters $\text{Amp}_i$ and $\sigma_\text{se}$ were selected ac-

according to the receptor spacing at the fovea (Coletta &

Williams, 1987; Dacey, 1993) and the physiologically

measured receptive field characteristics at the corre-

sponding region of the primate retina (Crone &

Kaplan, 1995). For simplicity, only the on-center,

off-surround-cells were considered. The membrane

potential of the $i$th sustained cell, $w_i$, is described by

with the output given by $y(t) = (\gamma/\delta) s(t) z(t)$, where $s$, $z$, $y$ represent the concentrations of $S$, $Z$, and $Y$, respectively and $\gamma$, $\delta$, $x$ denote rates of complex formation, decay to inactive state, and dissociation, respectively.

This equation has been used in a variety of neural
Stage 3: Quadratic-non-linearity-with-threshold and persistence. The “membrane potential” of the ith cell is transformed into an output signal (e.g., spike frequency) through a quadratic non-linearity with threshold, \( \lambda([w_i - \Gamma_s]_+)^2 \), where \([a]_+\) denotes the threshold, or half-wave rectification, function (i.e. \([a]_+ = a\) if \(a > 0\) and \([a]_+ = 0\), otherwise). Parameters \( \lambda, \Gamma_s \) represent the gain and the threshold level of this function, respectively. The thresholded signal provides the input to the additive equation

\[
\frac{dv_i}{dt} = \sigma(-v_i + \lambda([w_i - \Gamma_s]_+)^2),
\]

whose parameter \( \sigma \) determines the overall temporal persistence of the signal in the parvocellular pathway.

A.2.2. Retinal cells with transient activities (magnocellular pathway)

For simplicity, in our previous work the equations used to represent the retinal cells with transient activities did not incorporate the cells’ spatial receptive field profile. In order to simulate the spatial spread of masking effects more accurately, we incorporated the spatial receptive field profile of transient cells using a Gaussian kernel whose parameters (see Table 2) reflect physiologically measured receptive field characteristics of the transient cells in the primate retina (Croner & Kaplan, 1995). The surround of the receptive-field integrates inputs with low sensitivity but over a relatively large retinal area. The relatively small one-dimensional stimuli used in our simulations (see Section A.5) do not produce any appreciable surround response. Therefore, we used only the center of the receptive field in a shunting equation given by:

\[
\frac{dy_i}{dt} = -A_i y_i + (B_i - y_i) \sum_{j=i-n_i}^{i+n_i} G_j \{I_j(t) - I_j(t - \delta)\}
\]

\( (A.8) \)

A delayed version of the input (delay = \( \delta \)) is subtracted from the input to generate transient responses (backward-difference formula).

A.3. The post-retinal network

Because of its staggering complexity, a detailed model of the post-retinal network (LGN and visual cortical areas) would be computationally intractable. Our approach is to use a lumped network that is tailored according to the requirements of the simulation. For example, a study using achromatic stimuli would not require a detailed modeling of areas processing chromatic aspects of stimuli. In Purushothaman et al. (2002), we studied the coding and discrimination of edges for a wide range of edge blur and stimulus contrast values. This required a post-retinal network consisting of multiple “spatial-frequency channels” and two distinct layers to produce the contrast normalization. On the other hand, our study of motion deblurring using small targets did not necessitate such complexity (Purushothaman et al., 1998). Here, we will adopt the relatively simple representation used in Purushothaman et al. (1998) with one major exception: We will introduce post-retinal sustained-on-transient inhibition which is critical for the experimental paradigm studied in this paper.

A.3.1. Post-retinal cells mainly driven by the parvocellular pathway (“post-retinal sustained cells”)

The activity of the ith cell, \( p_i \), is given by the shunting equation

\[
\frac{1}{\tau} \frac{dp_i}{d\tau} = -A_p p_i + (B_p - p_i) \left\{ \Phi(p_i) + 2v_i(t - \eta) \right\} - p_i \left\{ \sum_{j=i-n_p}^{i+n_p} \Phi(p_j) + \sum_{j=i-n_p}^{i+n_p} H^p_{j,i} v_j(t - \eta - \kappa_p) + \sum_{j=i-n_p}^{i+n_p} Q^p_{j,i} m_j \right\},
\]

\( (A.9) \)

where the excitation consists of the afferent parvocellular signal and a feedback signal. The inhibitory signal consists of feedback, feed-forward, inter-channel terms. Excitatory and inhibitory recurrent (feedback, re-entrant) signals are carried out through the non-linear function \( \Phi(x) = 10a(\{a + 1\}^2 - 1) \), if \( a < 0.05 \) and \( \Phi(x) = a(a + 0.975) \), otherwise. This function and its parameters were chosen to achieve sharpening of boundary signals for dynamic inputs (Ogmen, 1993). The inhibitory kernels \( H^p_{i,j} \) and \( Q^p_{i,j} \) determine the spatial spread of intra- and inter-channel inhibition, respectively. Parameter \( \eta \) represents the relative delay between the parvocellular and magnocellular signals. Parameter \( \kappa_p \) reflects the relative delay of the inter-channel inhibitory signal with respect to the excitatory signal.

A.3.2. Post-retinal inhibitory inter-neurons

The post-retinal inhibitory inter-neurons carry the inhibition from sustained cortical cells to transient cortical cells via the additive equation:
\[
\frac{dq_i}{dt} = -A_q q_i + B_q p_t,
\]

where \(q_i\) is the activity of the \(i\)th post-retinal inhibitory inter-neuron.

A.3.3. Post-retinal cells mainly driven by the magnocellular pathway ("post-retinal transient cells")

The post-retinal transient cells receive excitatory and inhibitory inputs from the magnocellular pathway and a post-retinal sustained-on-transient inhibition via the kernel \(Q^e\) yielding the shunting equation:

\[
\frac{dm_i}{dt} = -A_m m_i + (B_m - m_i)2[y(t)]^+ + m_i \left\{ \sum_{j \neq i} H_{mj}^i [y(t - \kappa_m)]^+ + \sum_{j \neq i} Q_{j1}^m q_j \right\},
\]

\[(A.11)\]

where \(m_i\) is the activity of the \(i\)th post-retinal transient cell. The function \([\cdot]^+\) denotes full-wave rectification that generates the "on-off" response characteristics of transient cells. Parameter \(\kappa_m\) reflects the relative delay of the intra-channel inhibitory signal with respect to the excitatory signal.

A.4. Parameter values

The equations used for sustained retinal ganglion cells are identical to those used in Purushothaman et al. (2002) and the corresponding parameter values can be found in Appendix A.1 therein. The post-retinal equations were modified from Purushothaman et al. (1998) and the corresponding parameters can be found in Appendix A.3. therein. The newly introduced or modified parameters are given in Tables 1 and 2.

### Table 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(\eta)</th>
<th>(\kappa_p)</th>
<th>(\kappa_m)</th>
<th>(A_q)</th>
<th>(B_q)</th>
<th>(A_n)</th>
<th>(B_n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>2</td>
<td>18</td>
<td>10</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Amp</th>
<th>(\text{tse})</th>
<th>(\text{pi})</th>
<th>(\text{mp})</th>
<th>(\text{mi})</th>
<th>(\text{pm})</th>
</tr>
</thead>
<tbody>
<tr>
<td>sd</td>
<td>6(60/23)</td>
<td>100</td>
<td>100</td>
<td>56</td>
<td>80</td>
</tr>
</tbody>
</table>

A.5. Simulation methods

The system of ordinary differential equations consisting of Equations (A.5)–(A.11) were solved numerically with the CVODE package. This package uses variable-coefficient forms of the Adams and backward differentiation formula methods (Cohen & Hindmarsh, 1994). The programs were written in C and were run on SUN workstations. Numerical solutions of large systems of ODEs can be very time consuming. The model was simplified to keep the simulations within reasonable bounds. The model contains only one spatial dimension, which was sampled at 500 positions (i.e. \(1 \leq i \leq 500\)). At the foveal inter-receptor spacing of 23 s, this results in a region of 3.2 deg extent. To simplify the computations, the convolution sums were carried out with a fixed extent given by \(n_t = 28\), \(n_i = 40\), and \(n_p = 120\). The target covered 19 spatial positions at 23 s spacing, this corresponds to a size of 7 min. It was flanked on both sides by masks of the same size. Center-to-center separation between the target and the mask was 40 spatial positions, corresponding to an edge-to-edge separation of 8 min. The magnitude of the target and the mask were 1 (arbitrary) unit above a background of 1 unit. The durations of the target and mask were 2 simulation-time units which correspond to \(2 \times 8\) ms = 16 ms. The "perceived brightness" of the stimuli was computed as the time-integrated activity of the post-retinal sustained cells responding to the target (computed at the 19 positions occupied by the target stimulus). In the reaction-time experiments, the observers had to report as fast as possible and as accurately as possible, at which of the two sides they perceived the target. First, we computed the differential (target-and-mask–mask-only) responses for the target and the mask. This differential response carries the critical information that would allow the observer to choose between the target-and-mask versus mask-only side. The reaction-time was computed as a weighted average of the latencies (computed as time to peak) for the transient and sustained responses:

\[
RT_{\text{contour-mask}} = \frac{\int T(\text{target}, \text{mask}) - \int T(\text{mask})] \text{lat}_T + [\int S(\text{target}, \text{mask}) - \int S(\text{mask})]}{\int T(\text{target}, \text{mask}) - \int T(\text{mask}) + [\int S(\text{target}, \text{mask}) - \int S(\text{mask})]}
\]

where \(\int T(\text{target}, \text{mask}), \int T(\text{mask})\) denote spatio-temporally integrated transient responses for target-and-mask and for mask only simulations, respectively. Similarly \(\int S(\text{target}, \text{mask}), \int S(\text{mask})\) denote spatio-temporally integrated sustained responses for target-and-mask and for mask only simulations, respectively. Symbols \(\text{lat}_T\) and \(\text{lat}_S\) represent the computed response latencies for transient and sustained responses. In the simulations, under normal conditions the magnitude of the transient responses was much higher than...
that of the sustained responses. This difference biased the weighted-average towards the faster response, as required by the experimental task. A reduction in the transient response implied a higher reliance on the sustained response, thereby causing an increase in the reaction time. As an example, the simulations for SOA = 16 ms produced $\int T(target, mask) = 123.898$, $\int T(mask) = 68.0266$, $\int S(target, mask) = 30.1125$, $\int S(mask) = 0$, $lat_T = 2$ and $lat_S = 15$ yielding the RT value of 6.55 shown in Fig. 3. In order to compare these values to ART values in Fig. 11 a linear scaling was used. We subtracted 7 to bring the baseline values approximately to 0 ms and multiplied by 55/7 to bring the peak ART to $\approx 55$ ms.

References


across the macaque visual system. *Journal of Neurophysiology*, 79, 3272–3278.


