

Attention Alters Feature Space in Motion Processing

Marc Zirnsak¹ and Fred H. Hamker^{1,2}

¹Department of Psychology, Institute II, Westfälische Wilhelms-University, 48149 Münster, Germany, and ²Department of Computer Science, Artificial Intelligence, Chemnitz University of Technology, 09107 Chemnitz, Germany

Attending to a particular feature modulates the activity of neurons throughout the visual field with the result that relevant features are enhanced while irrelevant features are suppressed. Do these modulatory influences merely lead to a gating of relevant features, or does attention have a direct impact on the representation of feature space, leading to a different percept depending on the content of attention? We observed that direction estimates of the static motion aftereffect drastically change when human observers attend to a stimulus whose motion direction differs from the one of the adaptor. This observation suggests that feature-based attention might operate by local magnifications of feature space between relevant and irrelevant features.

Introduction

Visual attention refers to our ability to focus on only a small part of the vast amount of information obtained from our visual environment. The neural basis of this ability relies on gain control, that is, mechanisms that alter the input–output ratio of individual neurons in visual brain areas. Attending to a particular location in visual space enhances neural responses to stimuli presented at this location, known as spatial attention (for review, see Reynolds and Chelazzi, 2004). Similar effects have been reported if attention is directed toward a specific, nonspatial attribute (Chelazzi et al., 1993, 1998; Motter, 1994a,b; Treue and Martínez Trujillo, 1999; Martínez-Trujillo and Treue, 2004; Bichot et al., 2005). This feature-based attention is proposed to operate globally throughout the whole visual field, that is, the activity of neurons is enhanced if their tuning characteristics match with the attended feature even if their receptive field location is distinct from the present focus of spatial attention. Moreover, these modulations are found whether the attended feature is currently present or not. For instance, Chelazzi et al. (1993, 1998) reported that the activity of inferior temporal cortex neurons was higher if the monkey had to remember a preferred stimulus than if it had to remember a less preferred one. Similarly, Treue and Martínez Trujillo (1999) demonstrated that the response of direction-selective neurons in the middle temporal area (area MT) to their preferred direction was enhanced, relative to a neutral condition, if the monkey attended to the same direction of motion that was presented outside the receptive field of the recorded neuron. Furthermore, the response to the preferred direction was suppressed if the monkey attended to the anti-preferred direction of the recorded neuron.

Related effects have also been observed in humans. In Saenz et al. (2002), the general design used by Treue and Martínez Trujillo (1999) was adapted and used in a functional magnetic resonance imaging study, replicating their findings in terms of the strength of the hemodynamic response. Furthermore, Boynton et al. (2006) demonstrated that the motion aftereffect measured at a stimulus location different from the attended stimulus was stronger if both stimuli shared the same than if they had the opposite direction of motion. Finally, in Saenz et al. (2003), subjects had to detect a change in speed of two spatially separated motion stimuli. Detection performance was significantly better when both stimuli moved in the same rather than in opposite directions.

While the above-described observations are in accordance with a global upregulation and downregulation of the neural representation of features according to the feature similarity gain model of attention (Treue and Martínez Trujillo, 1999), it remains unknown whether feature-based attention has a direct impact on feature space. That is, does feature-based attention merely scale the population response, resulting in a more active representation of relevant features without altering their identity, or does it in addition alter the encoded feature at the level of neural populations?

Materials and Methods

If feature-based attention indeed alters feature space, the static motion aftereffect (SMAE) caused by an unattended motion stimulus will change when attention is directed to a different direction of motion (Fig. 1A). The SMAE is the consequence of viewing a moving pattern for a prolonged period and designates the experience of motion while observing a stationary pattern. Thereby, the perceived direction of the stationary pattern points in the opposite direction of the moving pattern, due to an unbalanced population response to the stationary pattern presumably caused by neural adaptation of motion-sensitive cells preferring the previously viewed direction of motion (for review, see Mather et al., 2008). We will use this relationship to infer the direction of motion that was encoded by the neural population. We hypothesize that attending to a certain direction of motion will alter the encoded motion direction of an additional dissimilar stimulus, resulting either in an attraction toward or a repulsion away from the attended direction. However, if both stimuli share the same direction, no change in the encoded direction of the

Received July 22, 2009; revised March 18, 2010; accepted March 29, 2010.

F.H.H. was supported by the German Science Foundation (DFG HA 2630/6-1), the European Commission (FP7-ICT: EYeshots), and the Federal Ministry of Education and Research (BMBF 01GW0653). We are grateful to Tirin Moore, Rufin VanRullen, and Nick Steinmetz for valuable comments on a previous version of the manuscript.

Correspondence should be addressed to Marc Zirnsak at his present address: Department of Neurobiology, Stanford University School of Medicine, Stanford, CA 94305. E-mail: mzirnsak@stanford.edu.

DOI:10.1523/JNEUROSCI.3543-09.2010

Copyright © 2010 the authors 0270-6474/10/306882-09\$15.00/0

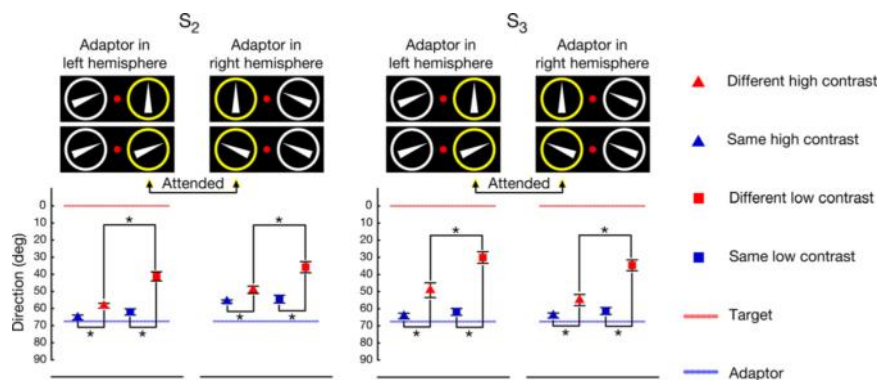


Figure 3. Contrast dependency of the SMAE. For both subjects the adaptor ($-67.5^\circ, 67.5^\circ$) was always directed inward with respect to the fovea. For the high-contrast condition, mean SMAEs of different-conditions are indicated by red triangles and mean SMAEs of same-conditions are indicated by blue triangles. For the low-contrast condition, mean SMAEs of different-conditions are indicated by red squares and mean SMAEs of same-conditions are indicated by blue squares. Dashed blue lines indicate the unsigned adaptor direction, and dashed red lines indicate the attended direction. Error bars denote 95% confidence intervals. For both subjects, differences between the same-condition and the different-condition are statistically significant in the high- and low-contrast conditions as indicated by asterisks. The low-contrast condition results in a significantly stronger attraction than the high-contrast condition.

contrast and a low-contrast condition. That is, we wondered whether the strength of the observed effect depends on stimulus contrast, as is suggested by electrophysiological experiments (Reynolds et al., 2000; Martínez-Trujillo and Treue, 2002). In these studies it was reported that the influence of spatial attention on the gain of single cells diminishes for high-contrast stimuli. If a related mechanism is working in the feature-based domain as well, we predict less pronounced distortions of the population response for high-contrast than for low-contrast stimuli and thus a reduced difference of the perceived aftereffect between the same- and different-conditions.

Subjects S_2 and S_3 participated in this control experiment. The general procedure and the task were identical to the main experiment, but here, both stimuli had a coherence of 100%. This was done to exclusively test the influence of stimulus contrast on the observed effect. Given equal coherence, a low-contrast stimulus should evoke weaker responses than a high-contrast stimulus. In the high-contrast condition, the dots of the target and the adaptor were white with a luminance of 85.6 cd/m^2 on a black background (2.2 cd/m^2), resulting in a Weber contrast of 37.91 for a single dot and for the whole RDK in a root mean square (RMS) contrast (Moulden et al., 1990; Seitz et al., 2006; Martínez-Trujillo and Treue, 2002) of 35.87. In the low-contrast condition, the dots of the target were also white with the same luminance as stated above, but the dots of the adaptor were gray with a luminance of 5.2 cd/m^2 , resulting in a Weber contrast of 1.36 and a RMS contrast of 1.29. Two directions ($-67.5^\circ, 67.5^\circ$) were used. Information about the size of the luminance change in the detection task is given in supplemental Table S4 (available at www.jneurosci.org as supplemental material). Approximately 20% of all the trials were excluded because the luminance detection performance of the subjects fell outside the predefined range (see above, Main experiment). The performance of the subjects is shown in supplemental Table S5 (available at www.jneurosci.org as supplemental material). Less than 0.02% of the remaining trials were identified and excluded as outliers, that is, the indicated direction fell outside the region of ± 2.5 SDs around the respective mean. The results are summarized in Figure 3 and supplemental Table S6 (available at www.jneurosci.org as supplemental material).

Control experiment 2: unattended motion. In this control experiment, we tested whether there is still a difference between the same-condition and the different-condition when attention is withdrawn from the target. This was done to exclude an interpretation of our results that is based on a simple weighting of both stimuli as would be predicted by the biased competition framework (Desimone and Duncan, 1995) if both stimuli were processed by the same population of neurons.

Subjects S_2 , S_3 , and S_4 participated in this control. Both the general procedure and the stimuli were identical to the main experiment. The

initial presentation of the to be attended target direction was replaced by a blank of the same duration (5 s). Instead of detecting a luminance change of the target, subjects looked for a luminance change of the central fixation point while ignoring both RDKs. The fixation point was white with a luminance of 85.6 cd/m^2 . The magnitude of the luminance change required central fixation, that is, it was assured that detection performance was at chance level if the subject fixated one of the two RDKs. The luminance change was 8.4 cd/m^2 for S_2 and S_3 , and 12.7 cd/m^2 for S_4 . The performance of the subjects is shown in supplemental Table S7 (available at www.jneurosci.org as supplemental material). Approximately 5% of all the trials were excluded because subjects indicated that they did not perceive an aftereffect. No outliers were detected. The results are summarized in Figure 4 and supplemental Table S8 (available at www.jneurosci.org as supplemental material).

Statistical analysis. Since the variability of the data is reasonably low ($SD = 5.57$, $SD_{SD} = 3.19$, $min = 1.27$, $max = 18.48$), ordinary statistics were used. It is assumed that all measurements are statistically independent. For all comparisons, Welch's generalization of the independent two-sample t test for unequal variances was used. All tests were conducted two sided. The test-wise α' level was adjusted due to Bonferroni correction so that $\alpha' = \alpha/m$, where m is the number of comparisons and $\alpha = 0.05$. The global null hypothesis is rejected if there exists at least one out of m multiple comparisons that is significant at the adjusted α' level.

In the main experiment, we tested separately for all subjects whether there was a statistically significant difference in the indicated SMAE directions between the same-condition and the different-condition and whether there was a difference between the same-condition and the baseline. This was done for all directions, consisting in each case of 10 comparisons for S_1 and S_2 ($\alpha' = 0.005$) and 2 comparisons for S_3 and S_4 ($\alpha' = 0.025$). Statistics are summarized in supplemental Table S3 (available at www.jneurosci.org as supplemental material).

In control experiment 1, tests were conducted to determine whether there was a difference between the same-condition and the different-condition for the high- and low-contrast conditions or between the different-conditions of the high- and the low-contrast conditions. In each case, two comparisons were made per subject. Statistics are summarized in supplemental Table S6 (available at www.jneurosci.org as supplemental material).

In control experiment 2, we tested whether there was a difference between the indicated SMAE directions of the same-condition and the different-condition. For each subject, two comparisons were made. Statistics are summarized in supplemental Table S8 (available at www.jneurosci.org as supplemental material).

Model. To explain the observed results, a "simple model" was set up in which the impact of attention on the population response was simulated. The model can be considered as a static, simplified version of earlier formalizations (Hamker, 2005, 2007). Thereby, the influences of attention were simulated in explicit terms. In general, feature-based attention is assumed to modulate the gain of individual neurons via feature specific feedback. This feedback is likely a combination of several signals originating at several stages in the cortical hierarchy ranging from high-level prefrontal to more low-level areas (Hamker, 2004).

Let $F = [0, 2\pi)$ be the feature space. Tuning curves are described by Gaussian functions located in F . For a given stimulus, the unmodulated or input activity r_i^{in} of the i th cell is as follows:

$$r_i^{\text{in}} = b_0 + b_1 \exp\left(-\frac{\phi_i^2}{2(\sigma_{\text{TC}})^2}\right)$$

$$\phi_i = ((p_s - c_i^{\text{TC}} + \pi) \bmod 2\pi) - \pi,$$

with $\phi \in [-\pi, \pi)$. The preferred direction, that is, the center of the tuning curve in feature space, is denoted by $c_i^{\text{TC}} \in F$. The direction of the stimulus is denoted by $p_s \in F$. The width of the tuning curve is determined by $\sigma_{\text{TC}} = 0.52$, $b_0 = 1$ is the baseline activity, and $b_1 = 10$ determines the height of the Gaussian. Note that cells have a bandwidth of 68° , which is the average bandwidth of motion selective cells as measured in V1 (Albright, 1984), as determined by the half-maximum width after subtraction of the baseline activity b_0 .

The influence of attention on the population response is formalized as a difference of Gaussians, as follows:

$$a_i = \exp\left(-\frac{\phi_i^2}{2(\sigma_A)^2}\right) - c \exp\left(-\frac{\phi_i^2}{2(3\sigma_A)^2}\right)$$

$$\phi_i = ((p_i - c_A + \pi) \bmod 2\pi) - \pi,$$

where $c_A \in F$ denotes the attended direction and p_i is the position of the i th cell in F . The extend of the central excitatory and the surrounding inhibitory region is determined by the constant c and by σ_A , which was set to $\sigma_A = 0.52$ for S_1 and $\sigma_A = 0.48$ for S_2 . Note that $c = 0.9$ for subject S_1 and $c = 0.8$ for subject S_2 . Further note that a simple Gaussian profile with central excitation and surrounding inhibition is inconsistent with the observed results, since it can account for attraction but not for the repulsion effect (see supplemental Fig. S2, available at www.jneurosci.org as supplemental material).

Finally, the modulated response r_i^{gain} of a given cell i is as follows:

$$r_i^{\text{gain}} = r_i^{\text{in}} (1 + w a_i),$$

where w is a weight that was determined by the data ($w = 4$ for subject S_1 and $w = 2.5$ for subject S_2). Note that if $r_i^{\text{gain}} < 0$, then it was set to $r_i^{\text{gain}} = 0$. For the simulations, the whole population consisted of 360 cells that were equally spaced in F .

With the above-described model, we are now able to simulate a population response for a given direction of motion for both the modulated and unmodulated conditions. To decode the direction represented by the population response, we computed the population vector of the neural activity in F as follows:

$$x = \sum_i r_i^{\text{gain}} \cos(c_i^{\text{TC}})$$

$$y = \sum_i r_i^{\text{gain}} \sin(c_i^{\text{TC}}).$$

The decoded direction is then given in Cartesian coordinates by the vector $\mathbf{v} = (x, y)$. For an overview of decoding methods, see, for example, Seung and Sompolinsky (1993) and Salinas and Abbott (1994).

We obtained the parameters by fitting the model to the data of subjects S_1 and S_2 to provide a qualitative explanation of the observed effect. Therefore, fitting was done initially by hand, followed by a more sophisticated exploration of a restricted part of the parameter space. Furthermore, to keep the description as simple as possible, we did not account for the apparent asymmetries of the data in the two hemispheres (see Fig. 2A).

For the illustration of the tuning curve shifts, a second population of cells was simulated. These cells pool the activity of the first population in feature space, that is, they are driven by a simple weighted sum of afferent responses (Rust et al., 2006). The bandwidth of the tuning curves of the second population is 83° , which is the average bandwidth of motion selective cells as measured in MT (Albright, 1984). The modulated ($w > 0$) and unmodulated ($w = 0$) tuning curves were normalized after the subtraction of the baseline, which was in this case defined as the minimal activity of a given tuning curve. Processing capacity is defined as the number of tuning curves n_{TC} for which the response to a given

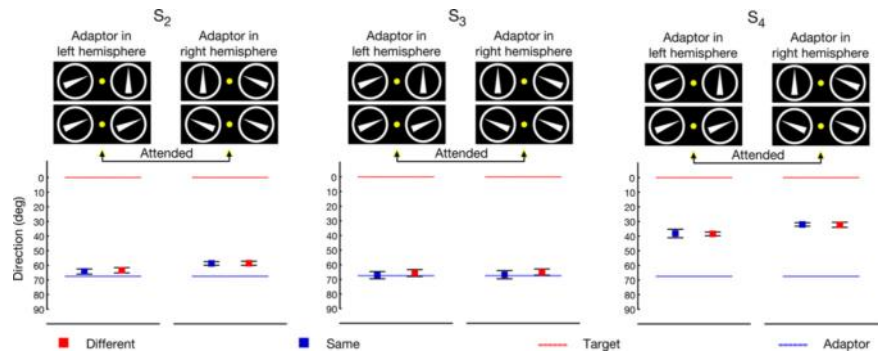


Figure 4. Results of the unattended motion condition. Subjects attended to the fixation point as indicated by the yellow color. For all subjects, the adaptor ($-67.5^\circ, 67.5^\circ$) was always directed inward with respect to the fovea. Mean SMAEs of different-conditions are indicated by red squares. Mean SMAEs of same-conditions are indicated by blue squares. Dashed blue lines indicate the unattended adaptor direction and dashed red lines indicate the attended direction. Error bars denote 95% confidence intervals. For all subjects, none of the differences between the same-condition and the different-condition are statistically significant.

direction exceeds half of the maximum activity of the respective cell [please refer also to Hamker et al. (2008) for the concept of processing capacity] and Δn_{TC} is the change in capacity in the modulated case with respect to the unmodulated one. For the reported simulations (Fig. 5F), the parameters of S_2 were used. Note that the same is true for the reported distortions of the population response (Fig. 5C,D).

Finally, as stated in the Introduction, feature-based gain modulations have been also reported during remembering previously attended features (Chelazzi et al., 1993, 1998). Thus, the model was extended to simulate additional gain modulations that might occur during the perception of the SMAE after the adaptation and thus would potentially be able to influence the perceived SMAE direction (see supplemental material, section 1.2, available at www.jneurosci.org). While this does not change the general assumption that feature-based attention is able to alter the represented direction of a motion stimulus already during the encoding, it has implications for the assumed gain profile. That is, simulating an additional gain modulation allows the “extended model” to be fitted to the data using a simple Gaussian modulatory signal consisting of central excitation and lateral inhibition as reported by Martinez-Trujillo and Treue (2004).

Results

We observe that attending to a peripherally presented direction of motion dynamically alters the direction of the SMAE, indicative of a change in the representation of the unattended stimulus in feature space. For a given adaptor, the perceived direction of the SMAE can deviate $>30^\circ$ from the expected direction. Figure 2 summarizes our main results. Note that the indicated SMAEs are transformed to the direction of motion that would have caused them, that is, the stimulus representation in the adaptation phase, assuming a simple reversal (Fig. 1A).

As is apparent in Figure 2, SMAEs of the same-condition often deviate from veridical but are often close to the baseline, that is, the perceived direction of the target when no additional task was required. When the motion direction of the adaptor differs from the one of the target (different-condition), however, attention to the target changes the encoded direction for the adaptor. Take subject S_1 for example (Fig. 2A): when the adaptor motion is to the left (-90°), attending a target with leftward motion indicates that the encoded adaptor motion is also leftward. The encoded motion of the same physical stimulus, however, drastically changes to approximately -60° when attention is directed to a target with upward motion.

Two patterns are observed, attraction and repulsion. For adaptor directions differing up to 90° from the target direction, the resulting SMAEs seem to arise from adaptor directions being

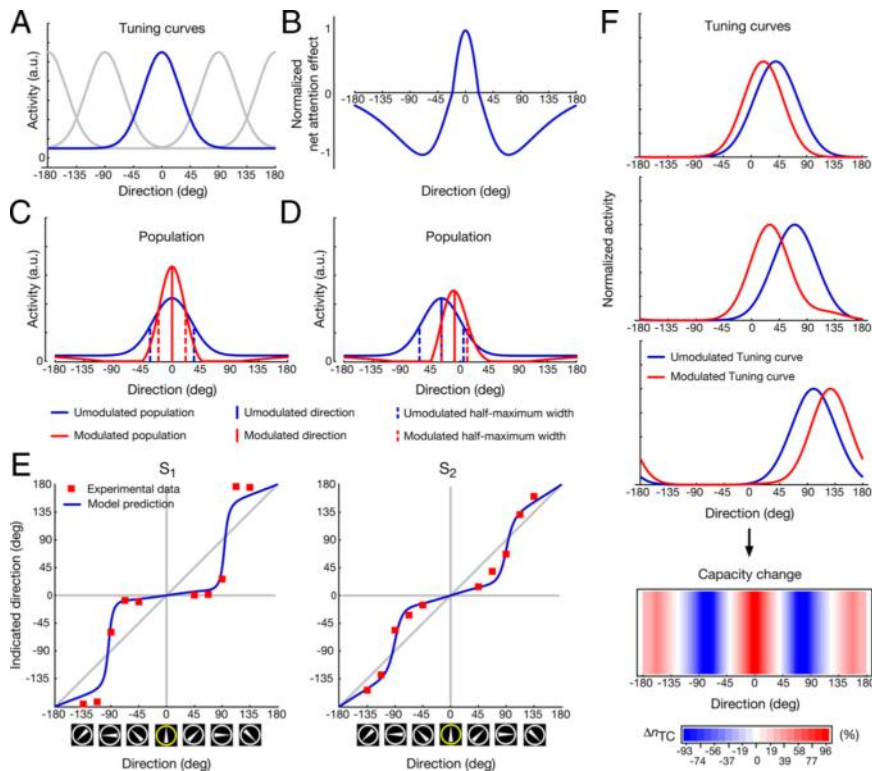


Figure 5. Simple model. **A**, Tuning curves of sensory neurons located on a circle in arbitrary units (a.u.). The bandwidth of the tuning curves is equal to the average bandwidth of motion selective cells measured in V1 (Albright, 1984). **B**, The net effect of attention. Directions close to the attended direction (0°) are enhanced, while directions farther away are suppressed. **C**, Consistent with experimental data (Martínez-Trujillo and Treue, 2004), the population response is sharpened as indicated by the half-maximum response when the stimulus direction is equal to the attended direction. **D**, For a stimulus with a direction of -30° , the population response is distorted, that is, the population vector of the neural activity is attracted toward the attended direction. **E**, Indicated SMAEs of subject S_1 and S_2 for 10 directions ($-135^\circ, -112.5^\circ, -90^\circ, -67.5^\circ, -45^\circ, 45^\circ, 67.5^\circ, 90^\circ, 112.5^\circ, 135^\circ$) denoted by the red squares shown together with the model fits (blue line) as a function of the veridical direction. Note again that SMAEs are shown as the direction that would have caused them, assuming a direct reversal. While directions near the attended direction (0°) are attracted, directions farther away are repelled. **F**, Exemplary tuning curve shifts of three model neurons. The center of the unmodulated tuning curve is $40^\circ, 70^\circ$, and 100° from top to bottom. While the center of the tuning curve shifts toward the attended feature for tuning curves with unmodulated centers close to the attended direction, it is shifted away for tuning curves with unmodulated centers farther away. This results in an increase of tuning curves that are processing the attended and the opposite direction. However, since the tuning curves with preferred direction farther away from the attended direction are suppressed, an effective increase in processing occurs at the attended direction, that is, tuning curves of neurons shift their preferred direction such that the whole population preferentially processes the attended properties.

closer to the target than they actually are, that is, they are attracted. This can be seen for the $-90^\circ, -45^\circ, 45^\circ$, and 90° adaptor directions in Figure 2*A* and for both adaptor directions (-67.5° and 67.5°) in Figure 2*B*. For adaptor directions farther away from the target, we find the opposite, that is, the adaptor directions are repelled with respect to the target, as can be seen for the -135° and the 135° adaptor direction in Figure 2*A*. Thus, the distance in feature space between attracted and repelled directions is increased.

Although there seems to be a trend of the measured baselines to lie below the SMAEs of the same-condition (Fig. 2*B*), they do not reach statistical significance (see Material and Methods for statistical details), with the exception of subject S_4 . The significant deviations for S_4 suggest that feature-based attention leads to a change of the encoded direction dependent on the actual neural representation rather than the direction that is inferred by the physical properties of the stimuli. Given the physical properties, one would not expect a difference between the baseline and the SMAEs of the same-condition. However, for S_4 the encoded direction does largely deviate from veridical if both motion stimuli

are unattended (Fig. 4). In the attended cases, they seem attracted toward the perceived direction of the attended target and are thus represented closer to veridical. Regardless, for all subjects all SMAEs of the different-condition do significantly deviate from their respective SMAEs of the same-condition.

It seems unlikely that a response bias can account for the observed pattern of attraction and repulsion. If there was a simple tendency of subjects to indicate the direction of the SMAEs as being closer to the attended target, we should not have found repulsion. Moreover, the magnitude of the effect inversely scales with the contrast of the adaptor as revealed by a control experiment in which we used high-contrast and low-contrast adaptors with 100% coherence (see Materials and Methods). For both the high- and the low-contrast conditions, we find significant deviations, but the magnitude in the high-contrast condition is significantly smaller than the one in the low-contrast condition (see Fig. 3 and supplemental Table S6, available at www.jneurosci.org as supplemental material). Although not yet tested in the feature-based domain, we interpret this contrast dependency as support for an attentional explanation since the magnitude of attentional effects has been previously reported to rely on stimulus contrast in the space-based domain (Reynolds et al., 2000; Martínez-Trujillo and Treue, 2002). Thus, if attention in both domains is characterized by related mechanisms at the single-cell level, we predict that gain modulations induced by feature-based attention will diminish for high-contrast stimuli.

Another potential caveat for an explanation based on feature-based attention would arise if both target and adaptor were simultaneously processed by a single population of cells with large receptive fields covering both stimuli. If this were true, attention would simply increase the contribution of the attended stimulus to the net SMAE (Alais and Blake, 1999) as known from the biased competition framework (Desimone and Duncan, 1995). While this again cannot explain the repulsion effect, it could be a possible explanation for the observed attraction. Thus, a second control experiment was conducted in which subjects had to detect a recurring luminance change of the fixation point and to ignore both target and adaptor (see Materials and Methods). Given the assumption that target and distractor are both processed simultaneously by the same cells, the biased competition framework predicts that SMAEs in the different-condition lie somewhere in between the SMAEs of the target and the adaptor. However, this is not what we observe. Consistent with a feature-based attention explanation, the data show that when subjects ignore the target, the effect disappears and no significant differences between the same- and the different-condition are observed (see Fig. 4 and supplemental Table S8, available at www.jneurosci.org as supple-

mental material). Thus, we conclude that both stimuli are not simultaneously processed by a single neural population with large receptive fields.

Finally, our results are also unlikely to result from eye movements as revealed by the analysis of the detection performance. The most reasonable case of breaking fixation for subjects would be to fixate the target to maximize the detection of the luminance change. However, the subjects performed much better in the experiments in which they were allowed to fixate the target than in the experiments in which they were required to keep central fixation (see supplemental material, section 1.3, available at www.jneurosci.org). Consistent with this, a subsequently conducted control in which eye movements were recorded revealed an average accuracy of 0.53° relative to the fixation point (see supplemental material, section 1.3, available at www.jneurosci.org).

Instead, our results can be explained by a class of models in which properties of the attended stimulus are fed back to early sensory areas and induce modulations of the population response. Initially we will focus exclusively on gain modulations during the adaptation phase (Fig. 5), which will be referred to as the “simple model.” While such attentional gain modulations relate to earlier models of feature-based attention (Hamker, 2006; Ardid et al., 2007; Reynolds and Heeger, 2009), feature-based distortions of the population responses have not been explicitly predicted, except for a study simulating a masking phenomenon known as feature inheritance (Hamker, 2007). The general idea of this model is that the gain resulting from feature-based attention is not uniform across the whole population, but rather peaks around the attended feature, which in turn distorts the population response such that it is effectively shifted toward the attended feature (Hamker, 2007). Our present model evaluation revealed that the net gain modulation ought to have a center-surround profile that follows a difference of Gaussians, that is, neurons that prefer directions close to the attended target are enhanced while neurons that prefer directions farther away are suppressed. Thus the strongest suppression is found for neurons preferring intermediate directions with respect to the attended one (Fig. 5B). Consistent with electrophysiological recordings (Martinez-Trujillo and Treue, 2004), this results in a sharpening of the population response if target and adaptor have the same direction (Fig. 5C) as determined by the half-maximum width of the response profile (see Material and Methods). For adaptors close to the target, however, the gain modulation leads to a distortion of the population in such a way that the population vector is shifted toward the target direction (Fig. 5D). Similarly, the vector of the population response to adaptors that lie in the suppressive surround is shifted away from the target direction. Altogether the distortions lead to local magnifications in feature space consistent with the response of the subjects (Fig. 5E). While these effects already occur due to simple gain changes of single neurons without shifts in the tuning curve, neurons that compute a weighted sum of the distorted population response ought to also alter their tuning properties, as predicted by the model (Fig. 5F). Tuning curves that are located in feature space close to the center of attention are attracted, whereas others located farther away are repelled. However, due to the suppressive effects surrounding the center of attention an effective increase in processing occurs at the attended direction. Thus, the whole population preferably processes the attended feature. Furthermore, such an increase in the number of neurons that are preferentially processing a common part of feature space has been theoretically shown to enhance coding accuracy (Abbott and Dayan, 1999; Shamir and Sompolinsky, 2006).

So far we have only considered the possibility of gain modulations during the adaptation phase, that is, the period during which both the adaptor and target RDK were presented. However, additional gain modulations could take place during the period during which the static RDK is shown and thus have a direct influence on the SMAE. Such modulations might be caused by the imagined target (Schlack and Albright, 2007) or in more general terms by a persistent representation of an already vanished stimulus (Chelazzi et al., 1993, 1998). Thus, we extended the simple model to simulate this possibility, that is, two subsequent gain modulations, one from feature-based attention during the adaptation phase and the other in the SMAE phase caused by the remembered target. Note that this “extended model” required us to explicitly simulate the static motion aftereffect (see supplemental material, section 1.2, available at www.jneurosci.org). As can be seen in supplemental Figure S4A (available at www.jneurosci.org as supplemental material), this model is consistent with the experimental observations as well. Furthermore, in line with Martinez-Trujillo and Treue (2004), the fit was achieved by assuming a simple Gaussian profile for both signals (supplemental Fig. S4E, G, available at www.jneurosci.org as supplemental material) instead of a difference of Gaussians as required for the simple model. However, the resulting gain profile is substantially smaller than expected from the data of Martinez-Trujillo and Treue (2004) (see supplemental material, section 1.4, available at www.jneurosci.org), which might be due to task differences as described in the Discussion. Furthermore, this model predicts a strong collapse of feature space during the encoding phase (supplemental Fig. S4B, available at www.jneurosci.org as supplemental material). While this prediction might be due to the specific assumptions made, for instance, the way we simulated the SMAE in the light of insufficient electrophysiological constraints, the collapse of feature space can be avoided with the extended model by assuming a difference-of-Gaussian profile for both modulatory signals (supplemental Fig. S4D, F, H, available at www.jneurosci.org as supplemental material). Thus, given the present uncertainties in the model assumptions, the extended explanation of our observations can be considered as an alternative interpretation.

Discussion

Attention is well known to integrate features, allowing for a coherent percept (Lu and Sperling, 1995; Treisman, 2006). Our study further reveals that in such an integration process, represented features are not fixed entities of the visual input, but rather dependent on the attentional state. Feature-based attention alters the neural representation of feature space as indicated by a change in direction of the SMAE dependent on the attended feature in the adaptation phase. These observations conceptually extend the feature-similarity gain model of attention (Treue and Martinez Trujillo, 1999). At the level of single cells, attention does still result in a simple scaling of the neural activity, which according to the feature-similarity model depends on a gain factor as a function of the similarity between the attended feature and the preferred feature of the particular neuron. However, since these changes in gain are not uniform, the modulations lead to distortions at the level of the population response, which changes the encoded feature. Thus, while previous experiments have only reported an upregulation or downregulation of the neural population response, which simply alters the strength of the encoded entity (Treue and Martínez Trujillo, 1999; Saenz et al., 2002; Liu et al., 2007), our observations suggest modulations of the population response governed by the feature-similarity principle, that

in addition enable the visual system to dynamically separate relevant from irrelevant features, that is, to increase the distance of the respective features in feature space. Furthermore, distorted population responses lead to tuning curve changes in neurons that are driven by these responses, for example, neurons that compute a weighted sum of the population response (Rust et al., 2006). This could explain the change in the spectral tuning of V4 neurons during a naturalistic visual search task, where cells became more sensitive to features of the attended search target (David et al., 2008). Thus, our model links physiological observations with perceptual experience.

As far as the origin of the distorted percept is concerned, it seems that a widely distributed network is involved in generating motion aftereffects (Taylor et al., 2000). However, as for the SMAE, both psychophysical (Moulden, 1980) and electrophysiological (Kohn and Movshon, 2003) evidence exist that it is already generated at the level of V1 (for review, see Mather et al., 2008); thus, distortions of the population response during the direction of feature-based attention should also be observable at this early processing stage. As a consequence, this early impact of feature-based attention will influence the representation of stimuli already before the final decision level has been reached. Although the distinction between encoding and decoding might be smoother than traditionally thought, earlier studies that focused on the discrimination of features attributed the misperception of stimuli to the level of decoding (Jazayeri and Movshon, 2007), for example, by optimally chosen top-down signals (Navalpakkam and Itti, 2007).

How does our work relate to the findings of Martinez-Trujillo and Treue (2004)? As already mentioned, their observed net effect can be described by a central excitation and lateral inhibition as well. However, their data suggest that lateral inhibition is strongest opposite to the attended direction, while according to our simple model, the peak inhibition is located much nearer to the attended direction in feature space and then gradually decreases moving farther away. While the extended model allowed us to explain our findings using a simple Gaussian signal, the estimated profile as constrained by our psychophysical data is substantially smaller than the estimated profile constrained by the electrophysiological measurements [see supplemental material, section 1.4 (available at www.jneurosci.org), for a fit of our model to the data of Martinez-Trujillo and Treue (2004)]. As outlined above, one explanation regarding these differences might be that our experiments presumably probe feature-based effects in V1, whereas Martinez-Trujillo and Treue (2004) recorded in area MT. Furthermore, the shape of the net attention effect does not necessarily have to be static but might change depending on the task at hand (Navalpakkam and Itti, 2007; Scolari and Serences, 2009). While subjects in our study were instructed to attend to the direction of the target RDK as accurate as possible, monkeys in the study of Martinez-Trujillo and Treue (2004) had to detect either a velocity change or a direction change of the attended stimulus. In the latter case, it could well be possible that the attentional system boosted neurons that were tuned away from the attended feature to facilitate the detection of another direction. Thus, the area of excitation in feature space should be rather broad as reported by Martinez-Trujillo and Treue (2004).

To summarize, according to the proposed mechanisms, the simple and the extended models are qualitatively equivalent in explaining our psychophysical findings of an altered feature space. While both models assume a major impact of feature-based attention in the adaptation phase, where the stimulus is

encoded, the extended model also predicts an influence of the remembered target in the SMAE phase, that is, a direct impact on our conscious percept. Thus, although it seems evident that top-down feature-based signals underlie our findings, the exact stage at which these signals act is a question for future research.

How do our results relate to findings in the spatial attention domain? Hamker et al. (2008) have demonstrated, using a computational model, that spatial attention directed to the saccade target explains the perisaccadic misperception of briefly flashed stimuli, known as compression of visual space (Morrone et al., 1997; Ross et al., 1997). In that model, misperception arises because spatial attention distorts the population response in the spatial domain, which in turn dynamically alters the receptive field profile. As a result, more cells are effectively processing the attended location. Indeed, a number of studies have indicated attentional changes of the receptive field structure in covert (Connor et al., 1996, 1997; Womelsdorf et al., 2006) and overt (Tolias et al., 2001) shifts of attention. Furthermore, although far less in magnitude, some perceptual effects have also been observed during covert attention shifts, such as repulsion (Suzuki and Cavanagh, 1997; Pratt and Turk-Browne, 2003) and attraction (Yamada et al., 2008). Similar to the effects reported in the present study, these different observations might be accounted for by a center-surround profile of the gain modulation (Hopf et al., 2006) as well. Thus, it seems that related mechanisms are involved in the feature and spatial domain of attention.

To conclude, attention has typically been characterized as a mere control of information flow, whereas the observations of perceptual distortions point toward an active recruitment of processing resources. As a result, a stimulus matching the content of attention is encoded more similarly to the attended one, whereas other stimuli can be repelled in feature space depending on the type of net modulatory profile as determined by the interactions at the systems level.

References

- Abbott LF, Dayan P (1999) The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11: 91–101.
- Alais D, Blake R (1999) Neural strength of visual attention gauged by motion adaptation. *Nat Neurosci* 2:1015–1018.
- Albright TD (1984) Direction and orientation selectivity of neurons in visual area MT of the macaque. *J Neurophysiol* 52:1106–1130.
- Ardid S, Wang XJ, Compte A (2007) An integrated microcircuit model of attentional processing in the neocortex. *J Neurosci* 27:8486–8495.
- Bichot NP, Rossi AF, Desimone R (2005) Parallel and serial neural mechanisms for visual search in macaque area V4. *Science* 308:529–534.
- Boynton GM, Ciaramitaro VM, Arman AC (2006) Effects of feature-based attention on the motion aftereffect at remote locations. *Vision Res* 46:2968–2976.
- Britten KH, Newsome WT (1998) Tuning bandwidths for near-threshold stimuli in area MT. *J Neurophysiol* 80:762–770.
- Chelazzi L, Miller EK, Duncan J, Desimone R (1993) A neural basis for visual search in inferior temporal cortex. *Nature* 363:345–347.
- Chelazzi L, Duncan J, Miller EK, Desimone R (1998) Responses of neurons in inferior temporal cortex during memory-guided visual search. *J Neurophysiol* 80:2918–2940.
- Connor CE, Gallant JL, Preddie DC, Van Essen DC (1996) Responses in area V4 depend on the spatial relationship between stimulus and attention. *J Neurophysiol* 75:1306–1308.
- Connor CE, Preddie DC, Gallant JL, Van Essen DC (1997) Spatial attention effects in macaque area V4. *J Neurosci* 17:3201–3214.
- David SV, Hayden BY, Mazer JA, Gallant JL (2008) Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. *Neuron* 59:509–521.
- Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193–222.

- Hamker FH (2004) A dynamic model of how feature cues guide spatial attention. *Vis Res* 44:501–521.
- Hamker FH (2005) The reentry hypothesis: the putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. *Cereb Cortex* 15:431–447.
- Hamker FH (2006) Modeling feature-based attention as an active top-down inference process. *BioSystems* 86:91–99.
- Hamker FH (2007) The mechanisms of feature inheritance as predicted by a systems-level model of visual attention and decision making. *Adv Cogn Psychol* 3:111–123.
- Hamker FH, Zirnsak M, Calow D, Lappe M (2008) The peri-saccadic perception of objects and space. *PLoS Comput Biol* 4:e31.
- Hopf JM, Boehler CN, Luck SJ, Tsotsos JK, Heinze HJ, Schoenfeld MA (2006) Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proc Natl Acad Sci U S A* 103:1053–1058.
- Jazayeri M, Movshon JA (2007) A new perceptual illusion reveals mechanisms of sensory decoding. *Nature* 446:912–915.
- Kohn A, Movshon JA (2003) Neuronal adaptation to visual motion in area MT of the macaque. *Neuron* 39:681–691.
- Liu T, Larsson J, Carrasco M (2007) Feature-based attention modulates orientation-selective responses in human visual cortex. *Neuron* 55:313–323.
- Lu Z-L, Sperling G (1995) Attention-generated apparent motion. *Nature* 377:237–239.
- Martinez-Trujillo JC, Treue S (2002) Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron* 35:365–370.
- Martinez-Trujillo JC, Treue S (2004) Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol* 14:744–751.
- Mather G, Pavan A, Campana G, Casco C (2008) The motion aftereffect reloaded. *Trends Cogn Sci* 12:481–487.
- Morrone MC, Ross J, Burr DC (1997) Apparent position of visual targets during real and simulated saccadic eye movements. *J Neurosci* 17:7941–7953.
- Motter BC (1994a) Neural correlates of attentive selection for color or luminance in extrastriate area V4. *J Neurosci* 14:2178–2189.
- Motter BC (1994b) Neural correlates of feature selective memory and pop-out in extrastriate area V4. *J Neurosci* 14:2190–2199.
- Moulden B (1980) After-effects and the integration of patterns of neural activity within a channel. *Philos Trans R Soc Lond B Biol Sci* 290:39–55.
- Moulden B, Kingdom F, Gatley LF (1990) The standard deviation of luminance as a metric for contrast in random-dot images. *Perception* 19:79–101.
- Navalpakkam V, Itti L (2007) Search goal tunes features optimally. *Neuron* 53:605–617.
- Pratt J, Turk-Browne NB (2003) The attentional repulsion effect in perception and action. *Exp Brain Res* 152:376–382.
- Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. *Annu Rev Neurosci* 27:611–647.
- Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61:168–185.
- Reynolds JH, Pasternak T, Desimone R (2000) Attention increases sensitivity of V4 neurons. *Neuron* 26:703–714.
- Ross J, Morrone MC, Burr DC (1997) Compression of visual space before saccades. *Nature* 386:598–601.
- Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. *Nat Neurosci* 9:1421–1431.
- Saenz M, Buracas GT, Boynton GM (2002) Global effects of feature-based attention in human visual cortex. *Nat Neurosci* 5:631–632.
- Saenz M, Buračas GT, Boynton GM (2003) Global feature-based attention for motion and color. *Vision Res* 43:629–637.
- Salinas E, Abbott LF (1994) Vector reconstructing from firing rates. *J Comput Neurosci* 1:89–107.
- Schlack A, Albright TD (2007) Remembering visual motion: neural correlates of associative plasticity and motion recall in cortical area MT. *Neuron* 53:881–890.
- Scolari M, Serences JT (2009) Adaptive allocation of attentional gain. *J Neurosci* 29:11933–11942.
- Seitz AR, Nanez JE Sr, Holloway SR, Watanabe T (2006) Perceptual learning of motion leads to faster flicker perception. *PLoS One* 1:e28.
- Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. *Proc Natl Acad Sci U S A* 90:10749–10753.
- Shamir M, Sompolinsky H (2006) Implications of neuronal diversity on population coding. *Neural Comput* 18: 1951–1986.
- Suzuki S, Cavanagh P (1997) Focussed attention distorts visual space: an attentional repulsion effect. *J Exp Psychol Hum Percept Perform* 23:443–463.
- Taylor JG, Schmitz N, Ziemons K, Grosse-Ruyken ML, Gruber O, Mueller-Gaertner HW, Shah NJ (2000) The network of brain areas involved in the motion aftereffect. *Neuroimage* 11:257–270.
- Tolias AS, Moore T, Smirnakis SM, Tehovnik EJ, Siapas AG, Schiller PH (2001) Eye movements modulate visual receptive fields of V4 neurons. *Neuron* 29:757–767.
- Treisman A (2006) How the deployment of attention determines what we see. *Vis Cogn* 14:411–443.
- Treue S, Martínez Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399:575–579.
- Womelsdorf T, Anton-Erxleben K, Pieper F, Treue S (2006) Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. *Nat Neurosci* 9:1156–1160.
- Yamada Y, Kawabe T, Miura K (2008) Mislocalization of a target toward subjective contours: attentional modulation of location signals. *Psychol Res* 72:273–280.

Supplemental Material

1.1 Modeling the contrast dependency of gain modulations

In the following we describe the proposed contrast dependency of feature-based attention. Note that the essence of this section can also be found in Hamker et al. (2008). In both the simple-model and the extended-model static neurons are used. Here the equation of gain modulation for static neurons from an equation used for dynamic neurons is derived. Let us assume we have a set of gain modulated neurons. The firing rate of each neuron can be described by a differential equation (Hamker, 2004, 2005a)

$$\tau \frac{d}{dt} r_i^{\text{gain}} = -r_i^{\text{gain}} + r_i^{\text{in}} + \left(A - \max_j (r_j^{\text{gain}}) \right)^+ w r_i^{\text{in}} a_i - w_{inh} r_i^{\text{gain}} \sum_j r_j^{\text{gain}}.$$

Such a gain function is motivated by several electrophysiological studies which have shown that feedback signals have a modulatory influence (e.g., Hupé et al., 2001; Moore and Armstrong, 2003) and it has successfully been applied to model the effect of feedback connections on feedforward processing (Hamker, 2005a). The term

$$\left(A - \max_j (r_j^{\text{gain}}) \right)^+$$

ensures that the efficiency of the feedback signal depends on the activity of the postsynaptic cell population, where $(\cdot)^+$ denotes half-rectification, meaning that if the argument is negative, it is set equal to zero. If the maximal firing rate exceeds the value A , the feedback signal a_i no longer affects the gain. This term has been shown to be consistent with a multiplicative contrast gain modulation as observed in several single cell recordings (Hamker, 2005b).

When we numerically compute the firing rate and set the weight of the dynamic inhibition among the cells to $w_{inh} = 0$, the change of activity Δr_i^{gain} in each time step is

$$\Delta r_i^{\text{gain}} \approx -r_i^{\text{gain}} + r_i^{\text{in}} + \left(A - \max_j (r_j^{\text{gain}}) \right)^+ w r_i^{\text{in}} a_i$$

When we ensure that $\max_j (r_j^{\text{gain}}) \leq A$ and further approximate

$$\frac{r_i^{\text{in}}}{\max_j (r_j^{\text{in}})} \approx \frac{r_i^{\text{gain}}}{\max_j (r_j^{\text{gain}})}$$

we obtain for the equilibrium a non-recursive equation for the firing rate of the gain modulated neurons:

$$r_i^{\text{gain}} = r_i^{\text{in}} \frac{1 + A w a_i}{1 + \max_j (r_j^{\text{in}}) w a_i} \quad (1)$$

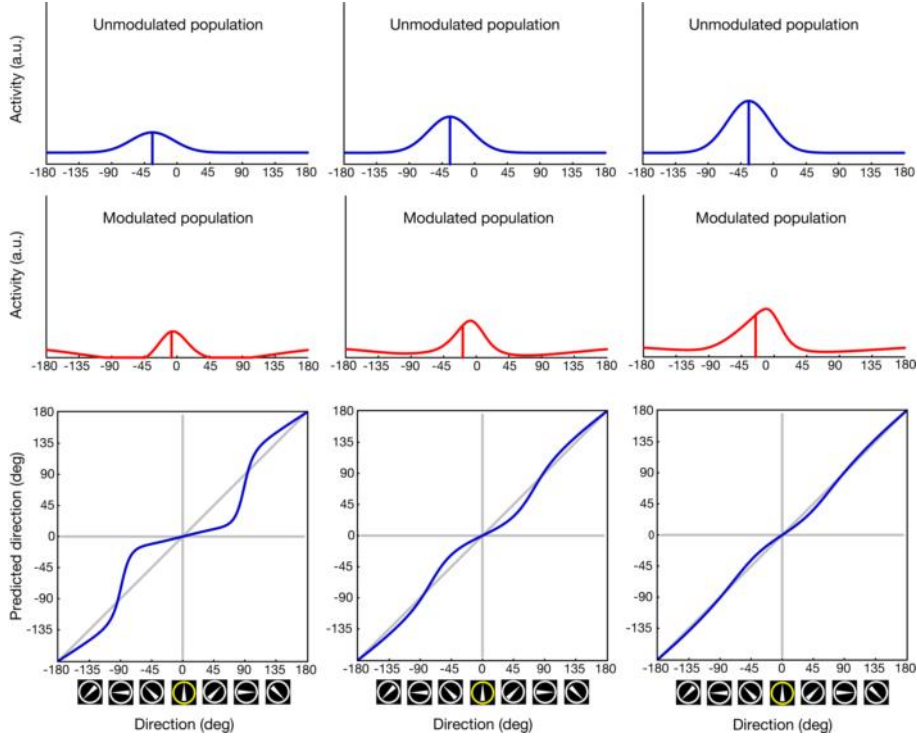


Figure S 1 Contrast dependency of gain modulations. From left to right the strength of the unmodulated population response in arbitrary units (a.u.) to a stimulus with a direction of -35° increases and thus the modulatory impact of attention decreases. Encoded directions are denoted by vertical lines. The respective unmodulated population response, the modulated population response, and the predicted distortions of feature space are shown from top to bottom.

that can further be simplified to

$$r_i^{\text{gain}} = r_i^{\text{in}} (1 + w a_i), \quad (2)$$

which was used in all described simulations.

To explain that the observed effect depends on stimulus contrast we can of course lower the weight w in Equation 2 in order to obtain a less distorted population response. However, a more plausible way would be to ensure that the strength of gain modulations does directly depend on the unmodulated input activity r^{in} as is true for Equation 1. That is, as is usually the case, stimuli of higher contrast lead to higher responses of cells preferring the stimulus properties. In analogy to the contrast dependency of gain modulations reported during covert shifts of space-based attention (Reynolds et al., 2000; Martinez-Trujillo, 2002) it is then assumed that the feature-based attention induced changes in activity diminish if the unmodulated response approximates its maximum A . Since Equation 1 only holds for

positive a_i values we have to apply the following generalization

$$r_i^{\text{gain}} = \begin{cases} r_i^{\text{in}} \left(2 - \frac{1 + A w |a_i|}{1 + \max_j (r_j^{\text{in}}) w |a_i|} \right) & \text{if } a_i \leq 0 \\ r_i^{\text{in}} \frac{1 + A w a_i}{1 + \max_j (r_j^{\text{in}}) w a_i} & \text{else} \end{cases}$$

This relationship between the initial activity and the magnitude of the modulatory influence, and thus the predicted mislocalization of stimuli in feature space is shown in Fig. S1. From left to right the input activity increases and the gain modulation vanishes.

1.2 Modeling the effect of gain modulations in the SMAE phase

Our simple-model of feature-based attention as described in the Materials and Methods section of the main article allowed us to explain the observed directional changes in the SMAEs assuming a modulatory profile that follows a difference of Gaussians. Assuming a modulatory attentional profile that simply follows a Gaussian profile as suggested by the experiments of Martinez-Trujillo and Treue (2004) does not allow to explain the observed repulsion effects. Fig. S2 shows the best fit of a model using such a Gaussian profile for both subjects S_1 and S_2 . It is apparent that the fit is very poor resulting in a proportional error reduction (*pre*) of only 12% for S_1 and 31% for S_2 as compared to 92% and 87% as obtained by the simple-model using a difference of Gaussians. The Gaussian modulatory signal can of course lead to a stronger distortion in the center where target and adaptor are similar, but this would result in a larger error when target and adaptor move in very distinct directions. Note, the proportional reduction in error measure is defined as

$$pre = 100 \frac{E_1 - E_2}{E_1}.$$

Thereby, E_1 is simply the sum of squared error (*SSE*) of the indicated direction with respect to the veridical (i.e., physical) direction and E_2 is the *SSE* of the indicated directions with respect to the corresponding model predictions.

However, in the simulations described so far it is assumed that the modulatory attentional impact on the population response does only occur during the adaptation phase. In the following we outline our simulations in which we assumed a second gain modulation during the SMAE phase where only the static adaptor was present.

In order to simulate a second gain phase directly modulating the population response which causes the SMAE we must first consider the latter. That is, how does the population response arise during the SMAE? We assume that the activity r'_i of cells during the SMAE is a linear function of the activity r_i of

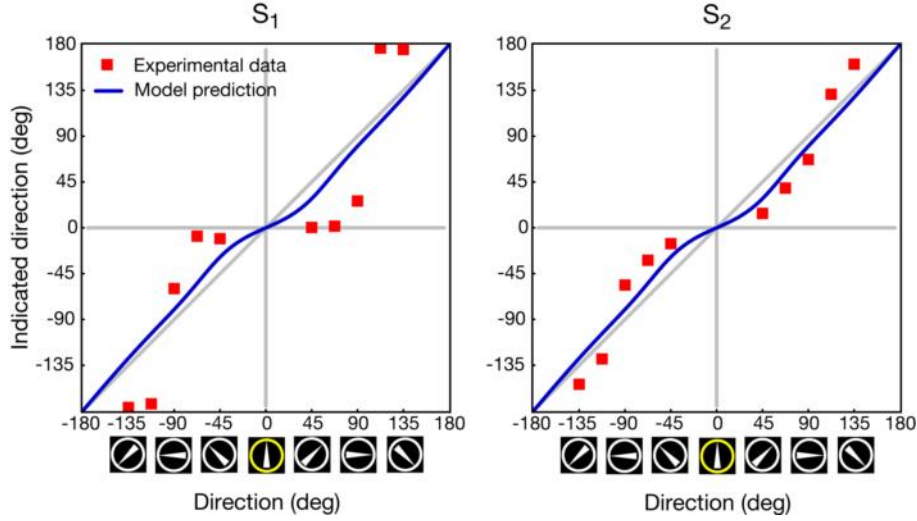


Figure S 2 Model fit using a Gaussian modulatory profile. Indicated SMAEs of subject S_1 and S_2 for all ten directions (-135° , -112.5° , -90° , -67.5° , -45° , 45° , 67.5° , 90° , 112.5° , 135°) denoted by the red squares shown together with the model fits denoted by the blue lines as a function of the veridical direction. As for the data, while directions near the attended direction (0°) are attracted directions farther away are repelled. However, a simple-model using a modulatory Gaussian profile predicts only attraction.

the same cell during the adaptation phase

$$r'_i = \frac{r_i b - r_i c_1 + c_2 c_1}{c_2},$$

where b denotes the baseline activity, c_1 denotes the maximal activity of the unmodulated case in the absence of any modulatory feedback, and c_2 the maximal activity of the modulated case. Further, r_i is equal to r_i^{gain} as defined in the Materials and Methods section of the main article which reduces to r_i^{in} in the absence of any feedback. For all reported simulations in this section we set $b = 1$, $c_1 = 11$, and $c_2 = 52.25$ which resulted in a range of r'_i from 1 to 11 which is the same as for the unmodulated response r_i^{in} used as input for the first gain stage.

The above described relationship is visualized in Fig. S3A. On the left three motion driven populations of the adaptation phase are shown. The corresponding static driven populations are shown on the right, where higher activity in the adaptation phase corresponds to lower activity in the SMAE phase. Note that each depicted population of the SMAE phase a decoding (Materials and Methods in main article) reveals exactly the opposite direction as compared to the direction obtained by decoding the populations of the adaptation phase. Thus, the assumed function of how neurons behave during the SMAE phase is consistent at the population level (e.g., Mather and Moulden, 1980) with the common observation that the SMAE is perceived in the opposite.

Fig. S3B contains a schematic sequence of the assumed gain modulations during the adaptation and SMAE phase. During the adaptation phase the population response caused by the adaptor RDK is modulated for the first time. This modulated population of the adaptor phase is then used as described

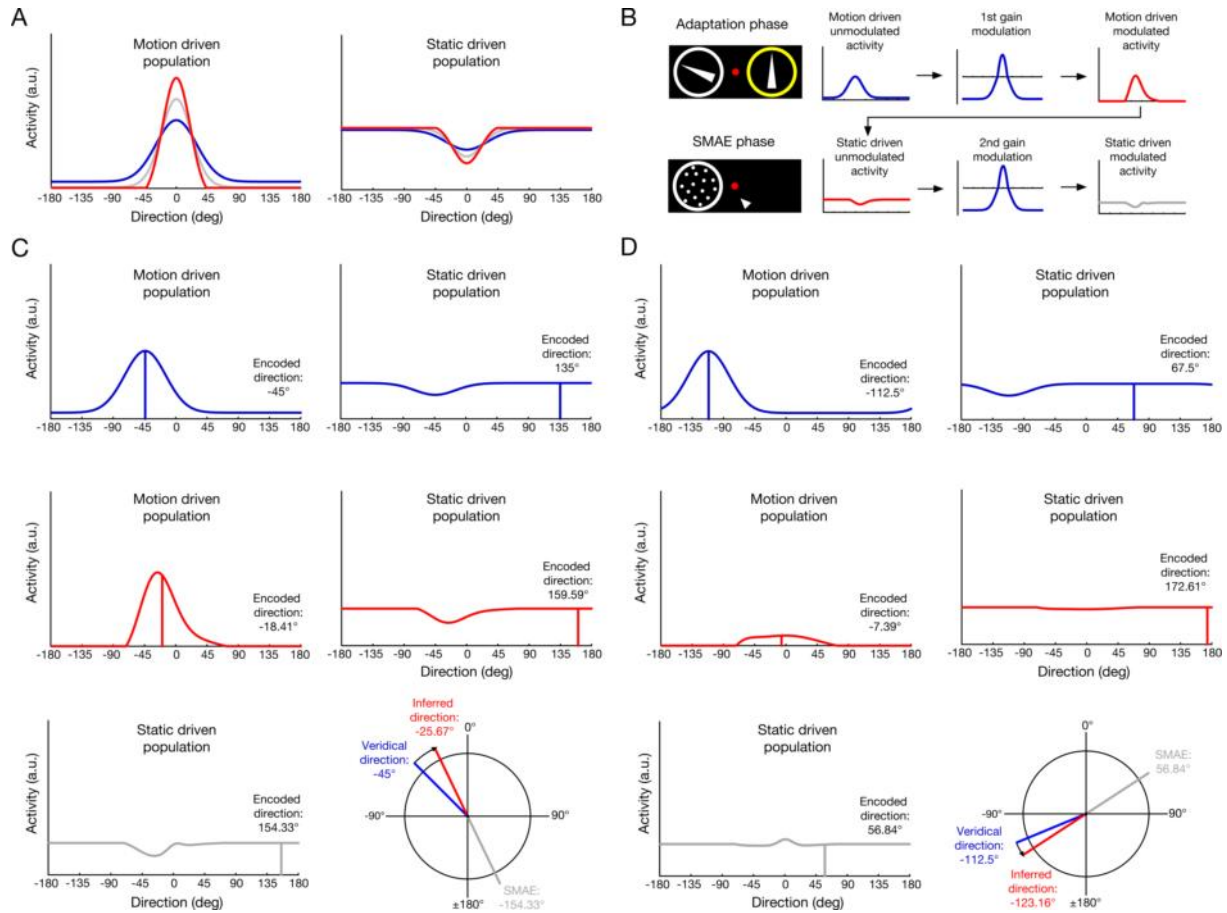


Figure S 3 Gain modulation during the SMAE phase. **A**, Assumed relationship between the motion driven population response and the static driven population response of the SMAE. **B**, Illustration of the temporal order and resulting populations of attention induced modulations during the adaptation and the SMAE phase. **C**, Illustration of attraction caused by the two subsequent gain modulations. **D**, Illustration of repulsion.

above to obtain the unmodulated population response of the SMAE phase on which the second gain modulation acts. The resulting modulated population response encodes the direction of the SMAE. Since the second gain modulation is assumed to be based on a remembered representation of the target RDK it is centered at the same position in feature space as the gain modulation during the adaptation phase. Fig. S3C shows an illustration in which the effect of both gain modulations result in an attraction of the encoded feature towards the center of attention in feature space. The blue motion driven population denotes the initially unmodulated response to a stimulus with a direction of -45° . If no distortions of the motion driven response occurs this will result in the corresponding static driven population (displayed to the right) which in turn results in a population that encodes a SMAE in the opposite direction (i.e., 135°). The red motion driven population denotes the distorted response due to the gain modulation in the adaptation phase. Again to the right the corresponding static driven population is shown. This population is then modulated by the second gain modulation resulting in the gray population response

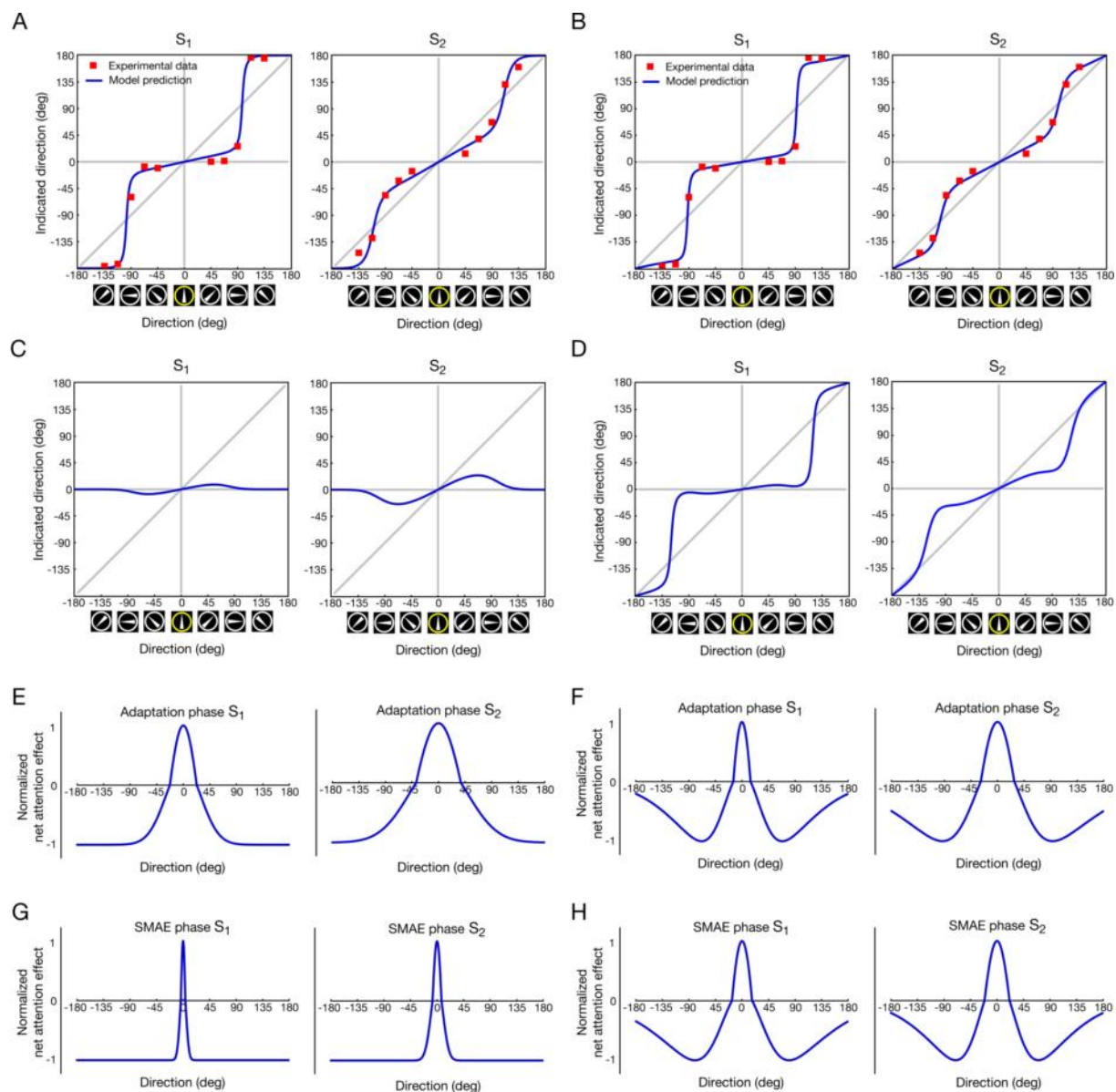


Figure S 4 Resulting fits and gain profiles obtained from the extended-model implementing two subsequent modulations during the adaptation and the SMAE phase. **A**, Fits for subjects S_1 and S_2 obtained by using simple Gaussian profiles. **B**, Fits for subjects S_1 and S_2 obtained by using difference of Gaussians. **C**, Predicted feature space during the adaptation phase by a model using Gaussian profiles. **D**, Predicted feature space during the adaptation phase by a model using difference of Gaussians as profiles. **E** and **F**, gain profiles of both models for the modulation during the adaptation phase. **G** and **H**, gain profiles of both models for the modulation during the SMAE phase.

which equals the SMAE. The direction of the SMAE is then reversed to obtain the inferred direction of -25.67° . Fig. S3D shows an example of repulsion with the same notation as in Fig. S3C.

To summarize, the extended-model using two subsequent gain modulation is in principle able to

explain both attraction and repulsion, by using only a simple Gaussian modulatory profile in feature space $F = [0, 2\pi)$

$$a_i = \exp\left(-\frac{\phi_i^2}{2(\sigma_A)^2}\right) - c$$

$$\phi_i = ((p_i - c_A + \pi) \bmod 2\pi) - \pi,$$

where $c_A \in F$ denotes the attended direction and p_i is the position of the i th cell in F . The extent of the central excitatory and the surrounding inhibitory region is determined by the constant c and by σ_A . This profile was used for all simulations for both the modulatory profile of the adaptation and the SMAE phase. As for the simulation reported in the Materials and Methods section of the main article we varied the width σ_A , the constant c , and the strength of the modulations to obtain a fit to our data. Only σ_A and the strength were varied independently for both gain modulation, whereas c was the same for both.

The quantitative fit of the extended-model lies in the same order of magnitude as the simple-model. The fits are visualized in Fig. S4 together with the resulting gain profiles. For S_1 we obtain a *pre* value of 96% as compared to 92% and for S_2 we obtain 85% as compared to 87%. The resulting gain profiles (Fig. S4G) of the SMAE phase are much smaller than compared to the gain profiles (Fig. S3E) of the adaptation phase which are again smaller than the gain profiles estimated for the feature-based attention effects in monkey MT (Fig. S5). Furthermore, this model predicts that feature space severely collapses during the adaptation phase (Fig. S4C). That is, all directions, including the opposite, are strongly attracted towards the center of the gain modulation. This effect can be reduced (Fig. S4D) by simulating both signals using a difference of Gaussians as described in the Materials and Methods section of the main article. For this model the goodness of fit in terms of *pre* is 96% for S_1 and 95% for S_2 , providing the best fit for the latter subject.

To conclude, while the above described extended-model allowed us to fit the data with about the same accuracy as the simple-model, it does not allow us to explain the broad gain profiles obtained from the data of Martinez-Trujillo and Treue (2004).

1.3 Eye movements

It is unlikely that eye movements caused the observed results of the main experiment and of control experiment 1. The most reasonable case of breaking central fixation for subjects would be to direct their gaze at the target to optimize the detection of the luminance change. However, if we compare the performance of subjects in the peripheral detection task of the main experiment and control experiment 1 to the performance of a central detection task, where subjects were instructed to fixate the target, it seems that subjects did not apply such a strategy as indicated by the performance in the respective conditions. If subjects had fixated the target during the peripheral detection task, we should have obtained a much higher detection rate as it is indicated by the performance of the central detection task.

The performance of the peripheral detection task is shown in Table S2 and Table S5. The mean performance across all subjects and the two experiments is 1.83 (*min* = 1.78, *max* = 1.87) as measured

in d' and corresponds to an unbiased proportion correct responses of $p(c)_{max} = 81.67\%$. For all subjects with the exception of S_4 the unbiased proportions of correct responses $p(c)_{max}$ are close to the empirical proportions of correct responses $p(c)$, indicating that response bias in the detection task was low. In the central detection task we find that for all subjects performance is close to perfect. For subject S_1 we obtain $d' = 3.66$, $p(c)_{max} = 96.66\%$, and $p(c) = 96.67\%$. For S_2 and S_3 we obtain $d' = 4.24$, $p(c)_{max} = 98.29\%$, and $p(c) = 100\%$. Finally, for S_4 we obtain $d' = 3.95$, $p(c)_{max} = 97.6\%$, and $p(c) = 98.33\%$. Note that for S_2 , S_3 , and S_4 $p(c)_{max}$ is less than $p(c)$. This is due to a common correction which was applied in order to estimate d' in the case of hit rates of $h = 1$ and false alarm rates of $f = 0$. For h the correction is $h = 1 - 1/N$ and for f the correction is $f = f/N$, where $N = 60$ is the number of detection trials (e.g., Macmillan and Creelman, 1991). For each subject the magnitude of the luminance change in this central task was set to the minimal luminance change which occurred for a specific subject in the peripheral task.

Furthermore, if subjects had fixated the target in the peripheral detection task, there would have been a foveal adaptation to the direction of the target. If we further assume that after the detection trials, where the target disappears and the adaptor stops moving, the subjects fixate the adaptor to indicate the SMAE, we will expect that in different-trials the indicated SMAE would always be directed opposite to 0° , that is, the SMAE should be 180° in the non-transformed description. As is obvious from Fig. 2 and Fig. 3, the effect is more gradual and such a pattern of eye movements would not have resulted in the observed repulsion effect, which should not have been found either if subjects had alternated between fixating the target and the adaptor.

Finally, the above reasoning is supported by a control which was conducted after the completion of the experiments reported in the main article. For each of the original subjects (S_1, \dots, S_4) we repeated a few trials of the main experiment while eye movements were recorded (*EyeLink II, SR Research, Ltd., Canada*). On average the accuracy of fixation was 0.53° relative to the fixation point with a standard deviation of 0.25° . Furthermore, no evidence of OKN-like eye movements were found in the direction of the target motion. Instead subjects tended to make small saccades (the vast majority being smaller than 1°) along the horizontal axis without a clear preference for a particular direction (i.e., left vs. right).

1.4 Modeling the feature-based attention effects as observed in monkey MT

Martinez-Trujillo and Treue (2004) conducted experiments in which monkeys had to attend to a certain direction of motion while the activity of neurons to another non-overlapping stimulus were recorded. By systematically varying the match between the preferred direction of the recorded neuron and the direction of the attended stimulus (the direction for the second stimulus was the same as for the attended) Martinez-Trujillo and Treue revealed a sharpening of tuning curves with respect to a neutral condition. Reconstructing the attentional modulation from these data led to the qualitative conception that feature-based attention follows a Gaussian modulatory profile with central excitation and lateral inhibition which is strongest opposite to the attended direction.

Here we provide a quantitative estimation of the modulatory profile by fitting our simple-model (see Materials and Methods of the main article) to the data of Martinez-Trujillo and Treue (2004). Fig.

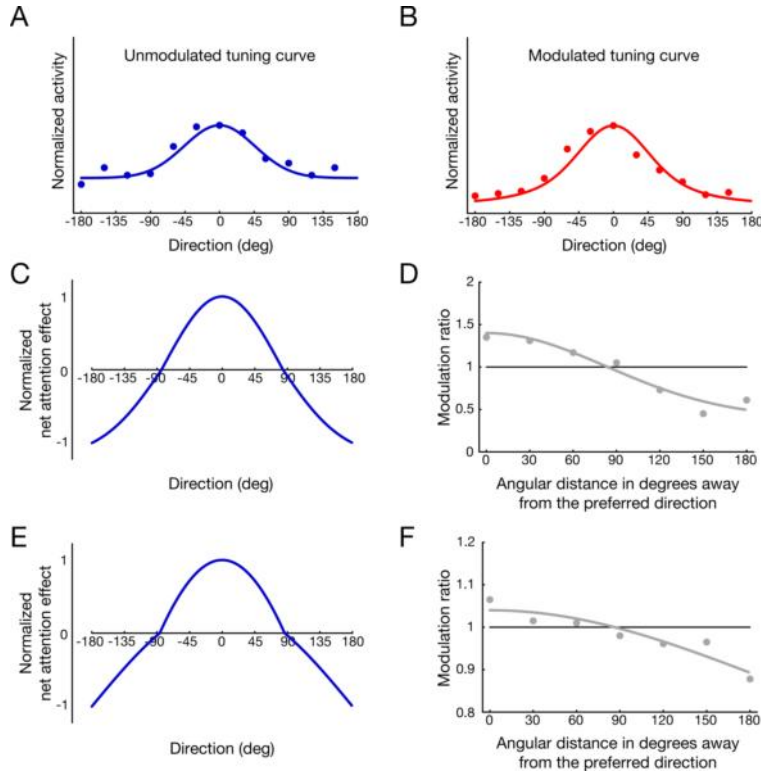


Figure S 5 The effect of feature-based in monkey MT as described by our model. **A**, Unmodulated tuning curve of a single MT cell (dots) shown together with the corresponding model curve (line). **B**, modulated tuning curve of the same cell shown together with the corresponding model prediction. **C**, Attentional modulation profile obtained by fitting the model to the modulation index of the cell. **D**, Modulation index of the single cell shown together with the model fit. **E**, Attentional modulation profile obtained by fitting the model to the modulation index of the population data. **F**, Modulation index of the population data shown together with the model fit.

S5 shows data of Martinez-Trujillo and Treue which were transcribed from the respective figures of the original article. Fig. S5A shows the unmodulated tuning curve of a single MT neuron (dots) (Fig. 4A, *attend-same* condition in the original article) together with the best fitting (least squares) model tuning curve (line). Fig. S5B shows the modulated tuning curve of the same neuron (Fig. 4A, *attend-fixation* condition in the original paper) together with the model fit which was obtained by minimizing the squared errors between model and data in Fig. S5D which designates the attentional modulation index for the above mentioned MT cell (Fig. 4B in the original article) used by Martinez-Trujillo and Treue to quantify their results. Fig. S5C shows the best fitting net attentional signal in our model. Fig. S5F shows the model fitted to the population data (Fig. 5 in the original article) and Fig. S5E the corresponding modulatory profile. According to their general appearance both the profile obtained for the single cell and the profile obtained for the population data are very similar (Fig. S5C and S5E) They consist of an excitatory area which extends to $\pm 90^\circ$ centered around zero and then becomes inhibitory in a monotonic manner with the strongest inhibition found for the opposite direction relative to zero.

Note in order to fit the model to the population data we used unmodulated tuning curves with small

(56°), medium (83°), and large (110°) bandwidths which corresponds to the average MT bandwidth \pm one standard deviation as reported in Albright (1984). We further varied the ratio of the peak activity to the baseline activity. The obtained modulatory profile does not depend on the exact tested parameter combination, that is, the above mentioned ratio and the bandwidth.

These simulation results confirm that the mechanisms of the model are in accordance to existing data further validating our model. Moreover, our analysis reveals that the task in the experiment of Martinez-Trujillo and Treue (2004) leads to a relatively broad attentional modulation across the space of motion direction.

References

- Hamker FH (2004) Predictions of a model of spatial attention using sum- and max-pooling functions. *Neurocomputing* 56C: 329-343.
- Hamker FH (2005a) The reentry hypothesis: the putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. *Cereb Cortex* 15: 431-447.
- Hamker FH (2005b) Modeling Attention: From computational neuroscience to computer vision. In L Paletta et al. (Ed.), *International workshop on attention and performance in computer vision (WAPCV 2004)* (pp. 118-132). Berlin, Heidelberg: Springer-Verlag.
- Hamker FH, Zirnsak M, Calow D, Lappe M (2008) The peri-saccadic perception of objects and space. *PLoS Comput Biol* 4: e31.
- Hupé JM, James AC, Girard P, Lomber SG, Payne BR, Bullier J (2001) Feedback connections act on the early part of the responses in monkey visual cortex. *J Neurophysiol* 85: 134-145.
- Macmillan NA, Creelman CD (1991) *Detection Theory: A User's Guide*. New York: Cambridge University Press.
- Martinez-Trujillo JC, Treue S (2002) Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron* 35: 365-370.
- Martinez-Trujillo JC, Treue S (2004) Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol* 14: 744-751.
- Mather G, Moulden B (1980) A simultaneous shift in apparent direction: further evidence for a "distribution-shift" model of direction coding. *Q J Exp Psychol* 32: 505-520.
- Moore T, Armstrong KM (2003) Selective gating of visual signals by microstimulation of frontal cortex. *Nature* 421: 370-373.
- Reynolds JH, Pasternak T, Desimone R (2000) Attention increases sensitivity of V4 neurons. *Neuron* 26: 703-714.

Table S 1. *Main experiment: luminance change.*

Subject	n	M	SD	min	max
S ₁	120	27.91	2.53	22.09	34.91
S ₂	111	28.2	1.78	23.36	31.66
S ₃	20	20.53	1.45	17.29	22.09
S ₄	33	32.63	5.57	21.27	41.08

Note. Unit of measurement is candela per square metre.

Table S 2. *Main experiment: performance.*

Subject	n	d'			$p(c)_{max}$			$p(c)$		
		M	SD	CI	M	SD	CI	M	SD	CI
S ₁	120	1.81	0.31	[1.76, 1.87]	81.49	4.03	[80.76, 82.21]	80.08	4.41	[79.30, 80.87]
S ₂	111	1.82	0.3	[1.76, 1.87]	81.53	3.93	[88.8, 82.26]	80.74	3.86	[80.02, 81.45]
S ₃	20	1.85	0.31	[1.71, 1.99]	82.02	3.95	[80.29, 83.75]	80	4.65	[77.96, 82.04]
S ₄	33	1.87	0.34	[1.76, 1.99]	82.23	4.21	[80.8, 83.67]	73.64	5.9	[71.62, 75.65]

Note. 95 % confidence intervals are shown. Unbiased proportion correct is denoted by $p(c)_{max}$. Proportion correct is denoted by $p(c)$. Both measures are shown in percent.

Table S 3. Main experiment: statistics.

S ₁										
Adaptor	Baseline		Same		Different		Baseline vs. Same		Same vs. Different	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t(df)</i>	<i>p</i>	<i>t(df)</i>	<i>p</i>
135	150.2	5.03	146.46	7.4	175.22	5.14	1.62(24)	.12	12.36(24)	3.88 · 10 ^{-12*}
112.5	130.62	12.2	119.89	16.34	176.54	2.82	2.05(25)	.05	13.23(14)	1.29 · 10 ^{-9*}
90	90.71	4.3	86.89	4.32	26.4	17.62	2.43(27)	.02	12.92(15)	9.1 · 10 ^{-10*}
67.5	51.3	8.05	45.92	12.77	1.59	5.17	1.38(23)	.18	1.47(18)	1.95 · 10 ^{-10*}
45	28.78	6.58	25.27	6.84	0.35	8.86	1.11(27)	.28	8.72(26)	3.03 · 10 ^{-9*}
-45	-29.34	4.93	-33.93	6.77	-10.58	6.57	2.13(25)	.04	9.59(27)	2.41 · 10 ^{-10*}
-67.5	-51.59	6.86	-43.84	10.8	-8.22	5.61	2.34(23)	.03	11.34(21)	2.02 · 10 ^{-10*}
-90	-91.78	7.01	-89.26	1.27	-59.65	18.48	1.37(14)	.19	6.19(14)	2.25 · 10 ^{-5*}
-112.5	-132.49	6.77	-129.66	8.21	-172.81	6.21	0.86(27)	.4	16.23(26)	3.82 · 10 ^{-15*}
-135	-148.77	5.76	-144.34	4.69	-176.38	8.16	2.31(26)	.03	13.18(22)	5.14 · 10 ^{-12*}

S ₂										
Adaptor	Baseline		Same		Different		Baseline vs. Same		Same vs. Different	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t(df)</i>	<i>p</i>	<i>t(df)</i>	<i>p</i>
135	140.07	5.63	134.94	3.66	160.72	3.77	2.95(24)	6.94 · 10 ⁻³	19.02(27)	1.54 · 10 ^{-17*}
112.5	117.77	4.65	116.19	3.58	131.18	3.98	1.05(26)	.3	10.85(27)	1.75 · 10 ^{-11*}
90	92.53	3.65	89.09	3.04	67.17	4.36	2.8(27)	9.21 · 10 ⁻³	15.97(25)	1.22 · 10 ^{-14*}
67.5	66.8	3.78	64.8	1.85	39.08	4.8	1.85(20)	.08	19.35(18)	1.58 · 10 ^{-13*}
45	42.58	2.49	39.54	4.19	14.13	5.04	2.42(22)	.02	15.02(27)	1.18 · 10 ^{-14*}
-45	-43.8	3.85	-40.5	2.47	-15.56	4.69	2.79(23)	.01	18.22(21)	1.95 · 10 ^{-14*}
-67.5	-66.68	4.24	-59.84	5.26	-31.83	6.37	2.83(26)	8.66 · 10 ⁻³	13.14(27)	3.10 ^{-13*}
-90	-91.51	2.27	-89.18	4.33	-56.23	3.9	1.84(21)	.08	21.88(27)	5.14 · 10 ^{-19*}
-112.5	-117.49	3.12	-116.39	2.18	-128.7	4.8	1.11(25)	.28	7.76(19)	3.84 · 10 ^{-7*}
-135	-139.91	6.28	-134.56	3.6	-153.6	4.48	2.85(22)	9.15 · 10 ⁻³	12.82(26)	6.13 · 10 ^{-13*}

S ₃										
Adaptor	Baseline		Same		Different		Baseline vs. Same		Same vs. Different	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t(df)</i>	<i>p</i>	<i>t(df)</i>	<i>p</i>
67.5	58.5	6.06	55.65	10.46	25.12	8.98	0.91(22)	.37	8.58(27)	3.05 · 10 ^{-9*}
-67.5	58.82	6.09	-53.33	10.67	-28.68	6.58	1.73(22)	.1	7.61(23)	9.28 · 10 ^{-8*}

S ₄										
Adaptor	Baseline		Same		Different		Baseline vs. Same		Same vs. Different	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t(df)</i>	<i>p</i>	<i>t(df)</i>	<i>p</i>
67.5	57.05	2.33	51.63	3.53	10.17	13.98	4.96(24)	4.4 · 10 ^{-5*}	11.14(15)	7.02 · 10 ^{-9*}
-67.5	-58.12	3.16	-50.79	3.93	-15.46	8.06	5.63(26)	5.74 · 10 ^{-6*}	15.27(20)	1.37 · 10 ^{-12*}

Note. Unit of measurement is degree. Number of measurements in each condition is $n = 15$. Critical p values are $p_{\text{crit}} = .005$ for S₁ and S₂, and $p_{\text{crit}} = .025$ for S₃ and S₄
* $p < p_{\text{crit}}$, two-tailed.

Table S 4. *Control experiment 1: luminance change.*

Subject	n	M	SD	min	max
S ₂	41	30.3	2.92	24.34	37.01
S ₃	41	17.96	2.55	10.67	22.09

Note. Unit of measurement is candela per square metre.

Table S 5. *Control experiment 1: performance.*

Subject	n	d'			$p(c)_{max}$			$p(c)$		
		M	SD	CI	M	SD	CI	M	SD	CI
S ₂	41	1.84	0.31	[1.74, 1.93]	81.78	4.03	[80.54, 83.01]	80	4	[78.78, 81.22]
S ₃	41	1.78	0.32	[1.68, 1.87]	80.99	4.12	[79.73, 82.25]	80.33	4.27	[79.02, 81.63]

Note. 95 % confidence intervals are shown. Unbiased proportion correct is denoted by $p(c)_{max}$. Proportion correct is denoted by $p(c)$. Both measures are shown in percent.

Table S 6. *Control experiment 1: statistics.*

S ₂							
Contrast	Adaptor	Same		Different		Same vs. Different	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t(df)</i>	<i>p</i>
High	67.5	65.27	2.87	58.4	2.77	6.68(27)	$3.05 \cdot 10^{-7*}$
High	-67.5	-55.87	1.74	-49.09	4.19	5.79(18)	$1.5 \cdot 10^{-5*}$
Low	67.5	62.0	3.65	41.21	5.24	12.61(24)	$2.43 \cdot 10^{-12*}$
Low	-67.5	-54.53	4.5	-35.84	6.44	9.21(25)	$1.62 \cdot 10^{-9*}$
Different vs. Different							
		Adaptor	<i>t(df)</i>			<i>p</i>	
		67.5	11.23(21)			$2.07 \cdot 10^{-10*}$	
		-67.5	6.68(24)			$6.57 \cdot 10^{-7*}$	
S ₃							
Contrast	Adaptor	Same		Different		Same vs. Different	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t(df)</i>	<i>p</i>
High	67.5	64.29	3.21	49.2	8.46	6.46(17)	$4.55 \cdot 10^{-6*}$
High	-67.5	-67.95	2.89	-54.95	6.41	4.96(19)	$8.06 \cdot 10^{-5*}$
Low	67.5	61.84	3.91	30.13	6.67	15.87(22)	$9.53 \cdot 10^{-14*}$
Low	-67.5	-61.33	4.01	-34.65	6.27	13.88(23)	$6.54 \cdot 10^{-13*}$
Different vs. Different							
		Adaptor	<i>t(df)</i>			<i>p</i>	
		67.5	6.85(26)			$2.55 \cdot 10^{-7*}$	
		-67.5	8.77(27)			$1.61 \cdot 10^{-9*}$	

Note. Unit of measurement is degree. Number of measurements in each condition is $n = 15$. Critical p value is $p_{\text{crit}} = .025$.
 $*p < p_{\text{crit}}$, two-tailed.

Table S 7. Control experiment 2: performance.

Subject	n	d'			$p(c)_{max}$			$p(c)$		
		M	SD	CI	M	SD	CI	M	SD	CI
S ₂	21	3.09	0.5	[2.88, 3.3]	93.33	3.14	[91.99, 94.67]	92.86	3.54	[91.34, 94.37]
S ₃	22	3.54	0.53	[3.32, 3.77]	95.68	2.36	[94.7, 96.67]	95.61	3.66	[94.08, 97.13]
S ₄	24	3.09	0.62	[2.84, 3.34]	93.02	4.08	[91.39, 94.66]	92.57	4.58	[90.74, 94.4]

Note. 95 % confidence intervals are shown. Unbiased proportion correct is denoted by $p(c)_{max}$. Proportion correct is denoted by $p(c)$. Both measures are shown in percent.

Table S 8. Control experiment 2: statistics.

S ₂						
Adaptor	Same		Different		Same vs. Different	
	M	SD	M	SD	$t(df)$	p
67.5	64.26	3.48	63.39	3.59	0.67(27)	.51
-67.5	-58.7	2.36	-58.59	2.74	0.12(27)	.91

S ₃						
Adaptor	Same		Different		Same vs. Different	
	M	SD	M	SD	$t(df)$	p
67.5	67.1	4.87	65.68	4.61	0.82(27)	.42
-67.5	-66.8	5.55	-64.99	4.23	1.01(26)	.32

S ₄						
Adaptor	Same		Different		Same vs. Different	
	M	SD	M	SD	$t(df)$	p
67.5	38.2	5.78	38.47	2.48	0.17(18)	.86
-67.5	-32.03	5.35	-32.29	3.45	0.24(23)	.81

Note. Unit of measurement is degree. Number of measurements in each condition is $n = 15$. Critical p value is $p_{crit} = .025$.