nature neuroscience

More is not always better: adaptive gain control explains dissociation between perception and action

Claudio Simoncini[1], Laurent U Perrinet[1], Anna Montagnini[1], Pascal Mamassian[2] & Guillaume S Masson[1]

Moving objects generate motion information at different scales, which are processed in the visual system with a bank of spatiotemporal frequency channels. It is not known how the brain pools this information to reconstruct object speed and whether this pooling is generic or adaptive; that is, dependent on the behavioral task. We used rich textured motion stimuli of varying bandwidths to decipher how the human visual motion system computes object speed in different behavioral contexts. We found that, although a simple visuomotor behavior such as short-latency ocular following responses takes advantage of the full distribution of motion signals, perceptual speed discrimination is impaired for stimuli with large bandwidths. Such opposite dependencies can be explained by an adaptive gain control mechanism in which the divisive normalization pool is adjusted to meet the different constraints of perception and action.

When we attend to objects moving in our environments, we can passively look at them and estimate all of their visual characteristics or we can actively track them with our eyes or arms. Do we use the same visual information or the same computations for solving these tasks? Two decades of active research comparing perception and action have attempted to determine whether the two systems have access to different information, leading to the two visual systems theory[1]. The outcomes are still unclear, and one cannot reject the alternative hypothesis that both perception and action rely on a generic visual processing[2]. Several studies have proposed a restricted comparison between two well-defined behaviors: motion perception and tracking eye movements (see ref. 3 for a review). These two behaviors present the advantages of sharing several of the key cortical stages of visual motion processing, particularly the pivotal medio-temporal (MT) area, which is present in both humans and monkeys. However, the results are conflicting. Although some early works indicated that perception and pursuit systems share similar performances when processing motion direction[4,5], others have reported different biases between the two types of responses[6,7]. By injecting background motion during either smooth pursuit or direction discrimination, two recent studies suggest that perception and action might be using the same information, albeit differently[8,9]. Indeed, one way to overcome the recurrent debate on the existence of a clear dissociation between ocular tracking and perception is to identify how the brain can decode visual motion information in an optimal way to fulfill the demands of both systems in a given context. It is therefore important to identify whether different or similar information processing might be involved using a single computational framework, as proposed for multisensory integration[10] or selective attention[11]. We used a different strategy and compared in human participants two low-level perceptual and motor tasks that require a decision on the same dimension: the object's speed. Moreover, we varied only the spatiotemporal richness of motion stimuli, keeping constant all other parameters, particularly their nominal speed. Thus, we were able to use a single computational framework to understand how humans process speed information for perception and action.

We introduced motion stimuli that consist of naturalistic textures with more complex features than sinusoidal luminance gratings while maintaining a tight control on their spatiotemporal structures[12,13]. Each stimulus contains multiple combinations of spatial and temporal frequencies lying along the same speed line in frequency space (**Fig. 1a** and **Supplementary Movies 1** and **2** for B_{sf} = 0.05 and 0.4, respectively). We varied the spread, or bandwidth (B_{sf}), of the spatial frequency distribution such that larger bandwidths correspond to richer stimuli while speed remains constant, as illustrated by the ellipse-like distribution of motion energy. Moreover, by setting both mean spatial (sf_0) and temporal frequency (tf_0), we were able to probe the contribution of different sets of spatiotemporal channels that sample motion from changes in luminance. We were therefore able to ensure that if differences between perception and action were found, they could not be a result of potential differences in their respective optimal speed or spatiotemporal frequency range. Overall, the experimental procedures for perception and action were kept very similar (**Fig. 1b**,**c**) in an attempt to minimize any possible contextual effects, such as perceptual learning or change in attention.

RESULTS

We first recorded ocular following in three participants. Ocular following involves reflexive responses, triggered at ultra-short latency in human and nonhuman primates, that exhibit many of the properties attributed to low-level motion processing[14,15,16] (see ref. 17 for a review). Initial eye acceleration of ocular following is known to be linearly related to input velocity for speeds below 80° s^{-1} (refs. 14,15,17). Moreover, monkey ocular following critically depends on extrastriate MT and medial superior temporal (MST) areas that are also

[1]Team InViBe, Institut de Neurosciences de la Timone, UMR 7289, CNRS and Aix-Marseille Université, Marseille, France. [2]Laboratoire Psychologie de la Perception, UMR8158, CNRS and Université Paris-Descartes, Paris, France. Correspondence should be addressed to P.M. (pascal.mamassian@parisdescartes.fr) or G.S.M. (guillaume.masson@univ-amu.fr).

Received 26 June; accepted 5 September; published online 30 September 2012; doi:10.1038/nn.3229

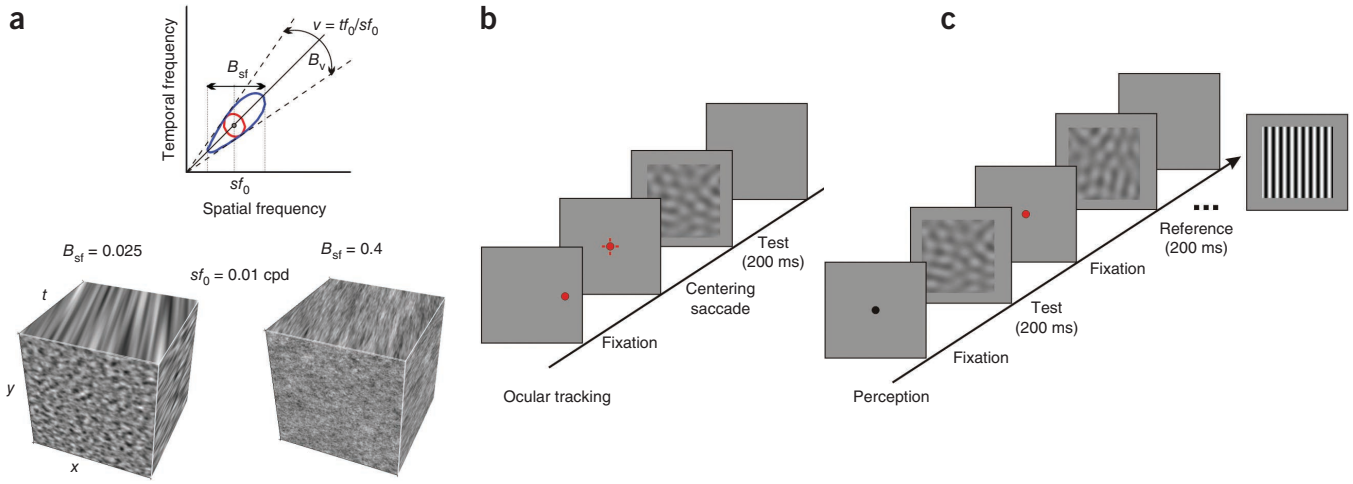

Figure 1 Band-pass motion stimuli for perception and action tasks. (**a**) In the space representing temporal against spatial frequency, each line going through the origin corresponds to stimuli moving at the same speed. A simple drifting grating is a single point in this space. Our moving texture stimuli had their energy distributed in an ellipse elongated along a given speed line, keeping constant the mean spatial and temporal frequencies. The spatio-temporal bandwidth was manipulated by co-varying B_{sf} and B_{tf}, as illustrated by the (*x*, *y*, *t*) examples. Human performance was measured for two different tasks, run in parallel blocks. (**b**) For ocular tracking, motion stimuli were presented for a short duration (200 ms) in the wake of a centering saccade[15] to control both attention and fixation states. (**c**) For speed discrimination, test and reference stimuli were presented successively for the same duration and participants were instructed to indicate whether they perceived the test stimulus as being slower or faster than the reference.

known to be pivotal in motion perception[18,19]. Lastly, speed tuning of these reflexive eye movements exhibit a marked similarity to that of both MST neurons and MT populations[17]. Thus, ocular following is an excellent probe for speed processing in human and nonhuman primates. Short-duration (200 ms) motion stimuli were presented in the wake of a 10° centering saccade, which was made to ensure that attention and gaze were always located at the same position[14]. All conditions were fully interleaved so that motion direction and bandwidth were unpredictable. **Figure 2a** illustrates, for a naive participant, mean eye velocity profiles of responses driven by an optimal (speed v = 20° s^{-1}, sf_0 = 0.3 cpd, contrast = 80%) rightward motion of a texture presented with different stimulus bandwidths B_{sf}. For all participants, increasing stimulus bandwidth resulted in stronger initial eye acceleration (for example, mean speed across participants: 1.5° s^{-1} with B_{sf} = 0.025, 6.1° s^{-1} with B_{sf} = 0.8) with little or no change in response latency (~90 ms). A log-linear relationship was found between response amplitude (mean speed in the 100–130-ms time window) and stimulus bandwidth (**Fig. 2b**). Moreover, response variability decreased when increasing stimulus bandwidth (mean s.d. across participants: 2.88° s^{-1} with B_{sf} = 0.025, 0.28° s^{-1} with B_{sf} = 0.8). Thus, higher stimulus bandwidths resulted in both stronger and more reliable reflexive pursuit initiations. To quantify these changes in reliability, we defined eye-movement sensitivity as the inverse of the s.d. of eye velocity during ocular following. Sensitivity increased with bandwidth for all three participants (**Fig. 2c**). Note that the eye-movement sensitivity is directly comparable to the perceptual sensitivity presented below. The same experiment was performed at four other mean spatial frequencies (sf_0 = 0.1–0.8 cpd), as well as for three different target speeds (20–80° s^{-1}) covering the complete range of both spatial frequency and speed tuning for human ocular following[15,17]. As expected[15], overall response amplitude varied with both speed and sf_0, but dependency on B_{sf} remained identical.

If larger stimulus bandwidths produce stronger and less variable responses, we can expect that ocular responses to two very nearby target speeds become more discriminable as bandwidth increases. We compared ocular following responses to textures translating at either 20 or 21° s^{-1} and presented at two different B_{sf} values. Responses were not distinguishable at small bandwidths, for all motion directions

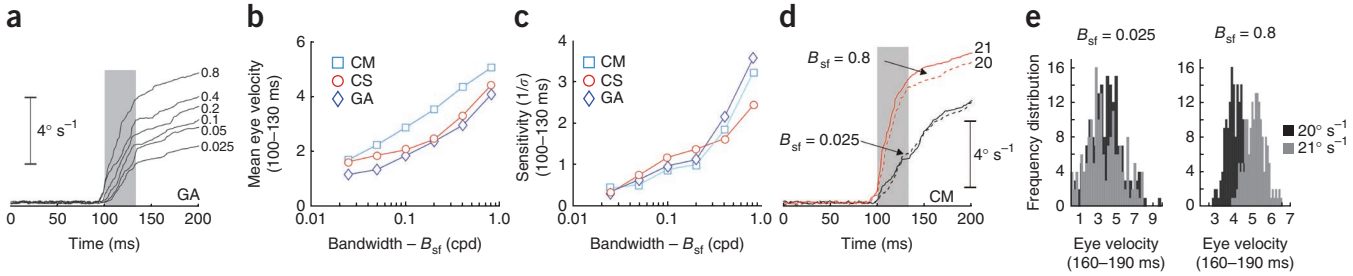

Figure 2 Ocular following responses to moving textures. (**a**) Mean (±s.e.m.) eye velocity profiles of ocular responses to moving textures of B_{tf} (0.025–0.8 cpd, indicated by the numbers to the right of each curve; naive participant GA). The gray shaded area illustrates the 100–130-ms time window used for quantitative analysis during the open-loop part of the response. (**b**,**c**) Mean (**b**) and sensitivity (s deg^{-1}; **c**) of initial eye velocity are plotted against B_{sf} for all three participants. (**d**) Comparison between ocular following responses elicited by moving textures drifting at either 20 or 21° s^{-1} and presented at two different bandwidths. (**e**) Distribution histograms of mean eye velocity during a late time window (160–190 ms) for the two B_{sf} conditions.

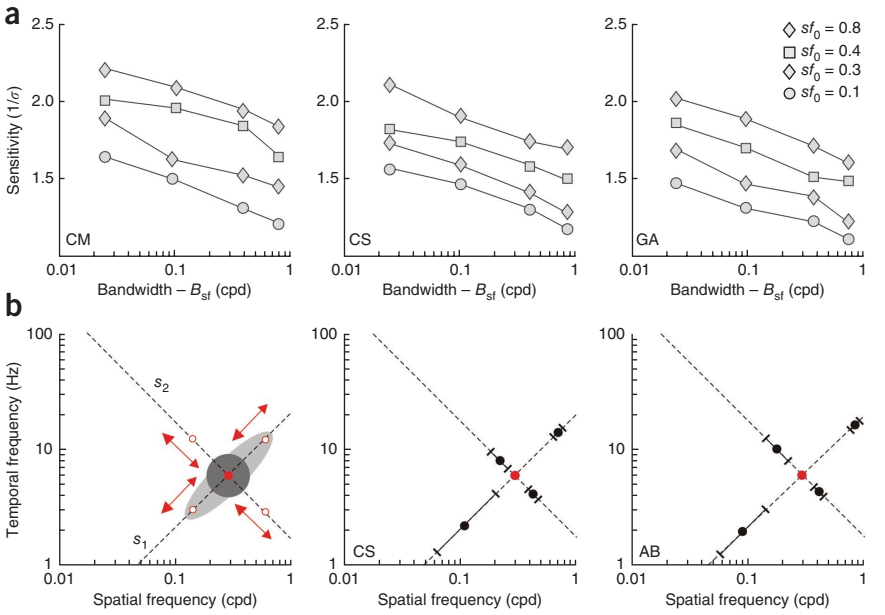

Figure 3 Effects of stimulus bandwidth on perceptual speed discrimination. (**a**) For each participant, sensitivity (inverse of s.d. used in the psychometric function) was plotted against stimulus bandwidth for four different mean spatial frequencies. (**b**) Left, ABX experiment in which a reference stimulus (open symbol) was compared with different test stimuli (closed symbols) distributed along the same speed line (dashed line s_1) or a line orthogonal (in linear space) to it (dashed line s_2). Dark and light gray areas plot the predicted distribution of discrimination performance for each condition, respectively. Middle and right plots, discrimination thresholds (±s.d.) measured in two observers (AB was naive).

and participants (**Fig. 2d**). On the contrary, responses clearly differed between the two target speeds when presented with high B_{sf}, at least for the later part of the open-loop tracking initiation. We plotted frequency distributions of mean eye velocity for all conditions (**Fig. 2e**) for the 100–130-ms time window. For the two participants, distributions were both wide and overlapping at low B_{sf}, but were narrow and clearly separated at high B_{sf}. We computed the sensitivity d' as a measure of discriminability between distributions for different speeds[20]. Sensitivity rose from 0.041 and 0.072 to 1.19 and 1.22 for participants CM and GA, respectively. With later time windows (for example, 160–190 ms), d' values further increased with high B_{sf} (1.56 and 1.47, respectively). We confirmed this effect at four other mean spatial frequencies, as well as with a lower contrast (20%, data not shown). Such an increase in oculometric sensitivity was also found when comparing much higher motion speeds (40 versus 42° s^{-1}). With the earliest time window (100–130 ms), mean (across participants) d' increased from 0.041 to 1.28 when varying B_{sf} from 0.025 to 0.8 cpd. Mean d' at later time window with broad B_{sf} further increased to 1.52, illustrating the temporal build-up of oculomotor sensitivity.

In a second series of experiments, the same three participants who took part in the original ocular following experiment were asked to perform a speed discrimination task[21,22] using the same set of motion stimuli. Observers had to compare two stimuli presented sequentially that varied in speed (**Fig. 1c**). Both stimuli were moving textures of the same mean spatial frequency (sf_0) and bandwidth (B_{sf}). One of the two stimuli (the standard) always moved at 20° s^{-1}, whereas the other (the test) moved slightly faster or slower because of a different mean temporal frequency. We measured the sensitivity to discriminate between speeds of two motion stimuli as a function of B_{sf} (**Fig. 3a**). Perceptual sensitivity was measured as the slope of the psychometric function to discriminate speeds and was taken as the inverse of the s.d. parameter of the best-fit cumulative Gaussian of the psychometric function. For all participants, sensitivity decreased as spatial frequency bandwidth increased. This result was in marked contrast with ocular following: participants were worse at discriminating two motion clouds when the motion textures were richer. Moreover, this effect remained constant over a large range of mean spatial frequency. Note that best speed sensitivity was observed for the highest mean spatial frequency (sf_0 = 0.8 cpd), which is one octave higher than the best spatial frequency for ocular following (0.3 cpd). Both bandwidth ($F_{4,62}$ = 5.27; P < 0.001) and spatial frequency ($F_{3,18}$ = 4.11; P < 0.001) effects were highly significant (repeated measures two-way ANOVA). Given that ocular following and perception might rely on different temporal integration, the same speed discrimination task was performed with six other stimulus durations, ranging from 100 ms to 1 s (data not shown). Increasing the stimulus duration resulted in overall higher sensitivities. However, sensitivity still linearly decreased with bandwidth for all stimulus durations. Thus, the opposite effect of bandwidth on speed discrimination and eye movements cannot be explained by different temporal integration for perception and action. Lastly, we asked whether varying bandwidth affected perceived speed in addition to speed sensitivity. We used a speed estimation task in which observers had to compare a moving cloud with a simple drifting grating. The grating was identical in all trials (sf_0 = 0.3 cpd, v = 20° s^{-1}). We confirmed the worsening of performance for richer textures: sensitivity decreased as the bandwidth of spatial frequency increased. Notably, changes in B_{sf} did not introduce any bias in perceived speed.

One potential caveat in the interpretation of the speed discrimination results is that participants could base their judgment on a particular combination of spatial and temporal frequencies content rather than on speed *per se*. For instance, the detrimental effect described above for speed discrimination might be a result of the broadening of the temporal frequency bandwidth co-occurring with the enlargement of the spatial frequency bandwidth. Others[22] have found that human motion discrimination relies predominantly on a speed-tuned mechanism rather than on the independent processing of temporal and spatial frequencies. To determine whether speed processing is the crucial mechanism involved with our stimuli as well, we decided to replicate the prior results using a matching-to-sample (ABX) task[23]. In each trial, participants had to discriminate between two stimuli, A and B, by matching one or the other with a third stimulus (X; X = A or X = B). For a given block of trials, stimulus A remained fixed at one mean spatiotemporal frequency and bandwidth, whereas stimulus B was varied in the spatiotemporal frequency space along either the diagonal of constant speed (condition 1) or a line orthogonal to it (condition 2; **Fig. 3b**). Note that the latter condition resulted in a change of the mean speed of stimulus B relative to stimulus A. The two conditions were run alternately, in random order. The same range of spatiotemporal frequencies (stimulus B) was tested in the two conditions. If participants compared speeds using any combination of spatial and temporal frequencies, their performance should

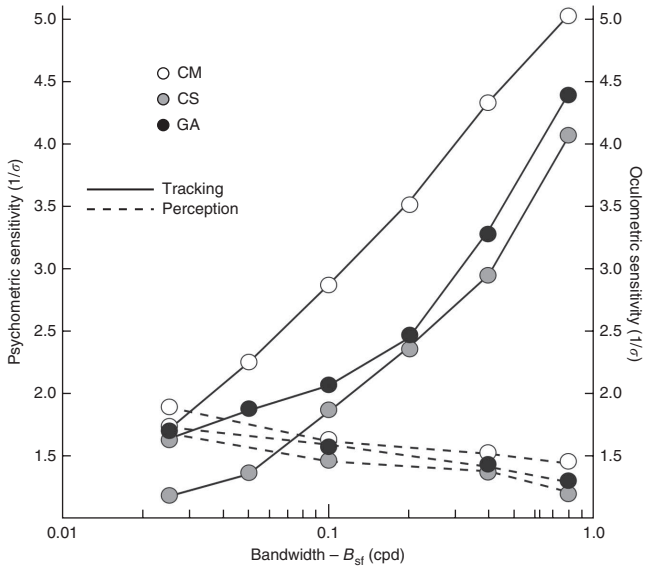

Figure 4 Comparing perception and eye movements. A direct comparison between perceptual (dashed lines) and oculomotor (solid lines) performances could be seen when plotting normalized sensitivities against B_{tf} for all three participants.

be identical between the two conditions, with ABX discrimination thresholds being distributed in a constant-radius circle. In contrast, if participants relied on a speed-tuned mechanism, their performance would be better in condition 2 than in condition 1: ABX thresholds would then be distributed on an elliptical contour whose small axis is along s_2. The ABX discrimination thresholds plotted for two participants followed the second prediction (**Fig. 3b**). They both displayed smaller thresholds (in terms of distance from the fixed stimulus A) along diagonal s_2 than along diagonal s_1. These results replicate the earlier finding[22] with our random texture movies. This pattern is the signature that perceptual discrimination of random moving textures is primarily based on speed information *per se* rather than on some other combination of spatial and temporal frequencies. These findings support the idea that the detrimental effect of increasing spatial bandwidth on speed discrimination is not a result of the broadening of temporal frequency bandwidth alone.

Given that both perceptual discrimination and ocular tracking performance are measured as speed sensitivity, we can directly compare their dependencies on the richness of the random moving textures. We plotted oculomotric and perceptual sensitivities against B_{sf} for each participant (**Fig. 4**). Clearly, perception and action exhibit opposite sensitivities to the richness of the motion stimulus, perception being penalized when provided with more spatiotemporal samples of the same pattern velocity, whereas ocular responses take full advantage of such supplementary and redundant information.

Contrast gain settings for perception and tracking

In all of the experiments presented above, the total energy of the random texture motion stimuli was kept constant across the different spatiotemporal distributions. The strong dissociation between perception and action (**Fig. 4**) could be explained by different gain setting mechanisms, and particularly by different normalization pools in spatiotemporal frequency space. In particular, a pure integration mechanism would result in higher gain control with increasing bandwidth[24,25]. On the contrary, a center-surround mechanism could result in a lower gain control for stimuli whose spatiotemporal frequency content is more spread around the mean frequency[25]. In a last series of experiments, we probed the gain setting mechanism of both responses while varying stimulus bandwidth. For both ocular following and speed discrimination, we presented to three participants, including two new naive ones, motion textures with different bandwidths and different total contrasts. We examined the mean eye velocity profiles of ocular following responses driven by a moving texture of either small (0.025 cpd) or large (0.8 cpd) B_{sf} (**Fig. 5a**). As contrast increased, responses became larger and decreased in latencies. However, ocular following was always stronger when presented with large bandwidth stimuli, consistent with the above results. In particular, the initial phases of eye velocity profiles appeared to be steeper, corresponding to stronger initial eye acceleration.

We plotted the mean response latency and earliest amplitude across participants (**Fig. 5b**). Eye movement latency decreased with contrast, as reported in earlier studies (see ref. 17 for a review), but the nature of this relationship was left unchanged when stimulus bandwidth

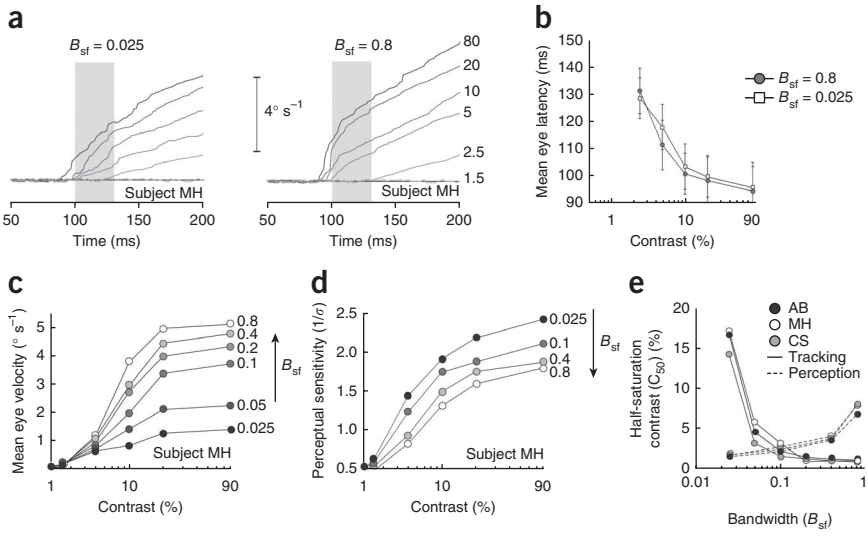

Figure 5 Contrast gain settings for perception and action. (**a**) Mean eye velocity profiles of ocular following responses to moving textures of increasing contrast (light to dark gray curves) and presented with small (left) or large (right) spatial frequency bandwidth (naive participant MH). (**b**) Mean (±s.d., across participants) response latency is plotted against pattern contrast for small (0.025 cpd) and large (0.8 cpd) spatial frequency bandwidths (open and closed symbols, respectively). (**c**) Mean eye velocity over the 100–130-ms time window is plotted against pattern contrast for increasing spatial frequency bandwidths, as indicated by the vertical arrow (naive participant MH). (**d**) For the same participant, perceptual sensitivity is plotted against pattern contrast, for each spatial frequency bandwidth. Increasing B_{sf} resulted in lower, shallower contrast responses, shifted toward higher contrast ranges. (**e**) Half-saturation contrast values obtained by fitting the contrast response functions for eye movements and perception with a Naka-Rushton function are plotted against spatial frequency bandwidth for three participants.

was varied. In comparison, contrast response functions of ocular responses recorded at different bandwidths differed sharply: response amplitudes decreased overall, but the dependency on contrast changed. We normalized the response amplitudes and fitted the normalized contrast response functions with the Naka-Rushton equation[26] (see Online Methods). For one participant, increasing B_{sf} shifted the best-fit function to lower contrast values (**Fig. 5c**). The slope of the contrast response functions ranged between 1.5 and 2 and was also affected, albeit more marginally. For each individual, best-fit half-saturation contrast was plotted against bandwidth. Small B_{sf} resulted in a higher half-saturation contrast (C_{50}) around 15–20%, consistent with previously reported values for single gratings[26]. Increasing B_{sf} rapidly decreased C_{50} to about 5%, as expected for a pure pooling mechanism[17]. Perception followed the exact opposite dependency, as would be expected from a divisive normalization mechanism, which has been reported with suppressive center-surround interactions[24]. Increasing stimulus bandwidth pushed the best-fit Naka-Rushton function toward higher contrast, marginally affecting its slope. Such reduction in contrast gain was illustrated by best-fit half-saturation contrast that nonlinearly increased with stimulus bandwidth in all three participants (**Fig. 5d**). Thus, broadening the spatial frequency content of the motion stimulus decreased contrast gain of perceptual responses, consistent with an inhibitory mechanism from spatiotemporal channels that are distant from the central frequencies.

Adaptive motion integration and divisive normalization

The opposite dependency on the bandwidth of spatiotemporal frequency information strongly suggests that the human visual motion system pools information differently for oculomotor and perceptual tasks (**Figs. 4** and **5**). Overall, the ocular tracking system takes full advantage of stimulus richness, achieving a much better estimate of target speed by pooling motion information across many different spatiotemporal channels. Initial tracking is often assumed to reflect a vector average computation performed over the whole population of speed-tuned units[19]. However, such a decoding strategy cannot explain why oculomotor sensitivity increased with stimulus richness. By increasing the spatial frequency bandwidth, more evidence of the same speed distribution is provided. Averaging multiple instances of the same information would give the same answer irrespective of the number of instances that are available. Clearly, the oculomotor system must use a decoding scheme that benefits from the additional evidence provided by larger bandwidths. In contrast, the perceptual system appears to be confused by this additional evidence.

A simple model can help us understand the difference between oculomotor and perceptual performance (see Online Methods). We reasoned that perception and action rely on similar encoding and decoding mechanisms, but that the opposite sensitivity to spatial frequency bandwidth results from a different pooling of information across speed channels. Such adaptive pooling may reflect the different constraints of information processing for either perception or action. Although perception requires that as much information as possible is preserved for solving a variety of different perceptual tasks (for example, speed discrimination and pattern discrimination), the control of eye movements requires an integration of information as quickly as possible for best estimating speed, disregarding other aspects such as the spatial structure of the visual scene. Several theoretical studies have suggested that gain control

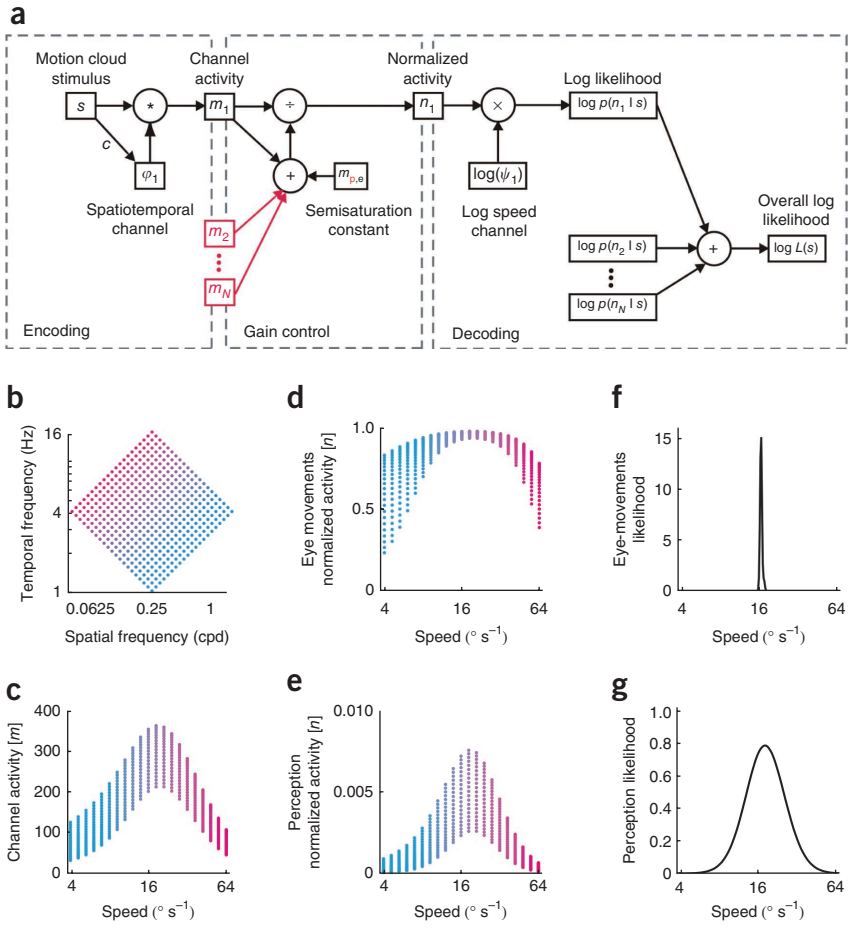

Figure 6 Model. (**a**) A schematic illustration of the models for eye-movements and perception. The models are composed of three stages: encoding of the stimulus, gain control and decoding of speed. Apart from the gain-control stage, the models are identical for the oculomotor and perceptual tasks. The model for perception includes the output of all channels in the normalization of the gain control stage (red elements). The asterisk in the encoding stage refers to the convolution operator. The properties of each step are further described in the subsequent subplots. (**b**) Tiling of spatiotemporal channels. Each channel is a bivariate normal function with its main axis oriented along a line of constant speed. All channels coding for the same speed are shown by the same color: blue for slow speed and pink for fast speed. The dots represent the center of the channels. (**c**) Channel activity. Each dot represents the activity of one channel for a stimulus moving at 20° s^{-1} with maximum contrast and maximum bandwidth. (**d**) Normalized activity for eye movements. In the eye-movement gain control, the normalization takes only the activity of the channel itself, so most channels almost reach their saturation level after normalization. (**e**) Normalized activity for perception. In the perception gain control, the normalization takes the summed activity across all channels, so only the most active channels remain active after normalization. (**f**) Eye-movement likelihood. The likelihood presents a narrow peak indicating that the model could precisely decode the stimulus speed. (**g**) Perception likelihood. The likelihood is now broadly distributed making it difficult for the model to discriminate two nearby speeds.

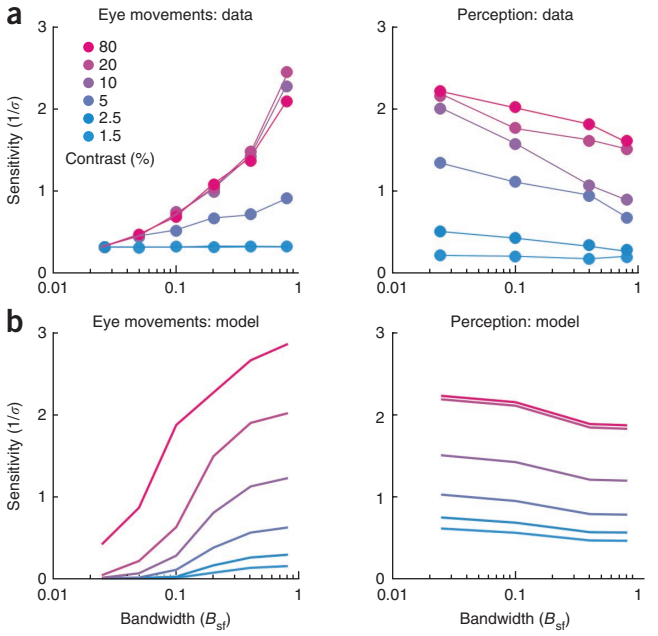

Figure 7 Comparison between experimental and model data. (**a**) Mean (across subjects) eye movements and perceptual sensitivities are plotted against spatial frequency bandwidth, for six different stimulus contrasts. As shown in **Figure 5**, increasing B_{sf} improved eye-movement sensitivity as well as contrast. On the contrary, perceptual sensitivity decreased with higher bandwidth. The reduction was even larger for high contrast. (**b**) Our model was fitted to these mean values for all conditions and both types of response simultaneously. Effects of both contrast and bandwidth were successfully simulated for our simple model. Best-fit values of the four parameters obtained with the complete data set are $\sigma[\log_2] = 2.11$, $\gamma = 6.58$, $m_p = 56.4$, and $m_p = 0.024$. A single change in the normalization pool of gain control is sufficient to explain the different changes in sensitivity of either perception or action, as well as the effects of contrast.

could be an efficient and universal mechanism for adaptive pooling of information across different channels[11]. Such a gain control is typically done by a generic scheme known as divisive normalization[24,25] and has been critically involved in many context-dependent sensory integration processes[10]. Thus, we tested whether gain control could also explain our results.

The encoding and decoding stages of our model are common to both perception and action (**Fig. 6a**). In the encoding stage, the motion cloud stimulus s is convolved with a bank of spatiotemporal channels φ_i, each tuned to a particular speed, with the property that they uniformly tile the spatiotemporal frequency space in log coordinates[22,27,28] (**Fig. 6b**). The tuning of the channels is modulated by the stimulus contrast c (ref. 28). In the gain-control stage, the activity of each channel m_i is normalized via a Naka-Rushton process with a semi-saturation constant m_e for the eye-movement and m_p for the perceptual task. The main difference in the two models lies in the suppressive fields of the normalization: each receptive field is restricted to the channel activity for eye movements and it is the summed activity across all channels for perception (**Fig. 6a**). Note, however, that more sophisticated gain controls, such as center-surround mechanisms, are of course possible and should be investigated in future studies[24]. In the decoding stage, estimated speed is extracted using the maximum-likelihood scheme previously proposed to model perceived motion direction[29,30]. The normalized activity n_i is multiplied by the logarithm of the speed channel ψ_i to produce the log-likelihood for this channel. These log-likelihoods are then summed to generate the overall log-likelihood of the stimulus (**Fig. 6c–g**).

We compared the model simultaneously to both eye movements and perception data to account for the effects of contrast and bandwidth on performance (**Fig. 7**). With only four parameters (see Online Methods), our model can reproduce most of the key features of the experimental findings. Increasing bandwidth increases the sensitivity of eye movement responses and decreases perceptual sensitivity (**Fig. 4**). Moreover, these two opposite effects are affected by the contrast of the stimulus in a way that is similar to the experimental findings. Increasing contrast results in improved sensitivity. However, larger spatial frequency bandwidth results in higher contrast sensitivity (that is, lower contrast gain) for eye movements, as illustrated by the wider change in sensitivity across the contrast range. The opposite result is obtained for perception with our model: change in sensitivity with contrast is larger at small bandwidth than at high bandwidth. Thus, keeping all of the model stages identical for both perception and action, but varying the size of the normalization pool, nicely reproduced the main features of our experimental results.

DISCUSSION

Comparing the performance of human action and perception is often blurred by the large experimental differences that exist between the two types of tasks, as well as by difficulties in finding common metrics. The original neuropsychological findings have led to the idea that different sensory processing mechanisms serve perception and action[1–3]. This view has been disputed and, from a large number of experimental studies, it has been argued that a single sensory processing system (for example, vision) can feed both perceptual and motor decision systems[31]. Our goal was to provide a solid experimental and theoretical framework that would resolve this long-lasting controversy. We found that ocular tracking initiation and motion perception present opposite sensitivities to the complexity of moving textures with natural-like statistics. By doing so, we were able to probe how the visual motion system pools information differently across several spatiotemporal frequency channels to extract a single entity, speed, but in different behavioral contexts. Our computational model strongly suggests that the property of a single nonlinear stage, namely gain control, can explain this pattern of opposite sensitivities. We found that a generic computation, divisive normalization[10,11], is essential for implementing such adaptive gain control by using different normalization pools.

We propose that both perceptual (speed discrimination) and motor (ocular tracking) responses have access to the same information but normalize it differently to match the different needs of perceptual and motor systems. Action often needs to summarize all of the information into a single value about location or speed. For the initiation of tracking eye movements, a rapid, linear integration of local motion information is dynamically modulated by nonlinear mechanisms[16,32]. A normalization stage operating in the same speed channel helps to improve the signal-to-noise ratio to quickly decode the velocity of the target to be pursued[17,19]. Different decoding schemes have been proposed in previous studies, such as vector averaging, maximum likelihood or winner take all[17,19,26,32]. Summing normalized motion information across different channels and using their likelihoods is the most efficient solution when decoding a population of macaque area MT neurons[30]. Our model suggests that estimating target speed for driving eye movements at very short latencies can be efficiently implemented by the same decoding computation, both in terms of initial eye acceleration and response reliability or sensitivity.

Such a decoding scheme was originally applied to perceptual detection and coarse discrimination of moving[29,30] and static patterns[33]. We found that we could successfully explain the sensitivity of human speed discrimination and its dependency on both spatial frequency bandwidth and contrast. The key difference with the processing underlying eye movements is the nature of the normalized population activity. When motion information is normalized using the complete set of spatiotemporal frequency channels, the normalized activity becomes more concentrated around its mean and the speed likelihood is consequently more widely distributed. Such a difference in the normalization pool size is sufficient to produce the opposite dependency on spatial frequency bandwidth that was observed for eye movements. It has been shown that such a divisive normalization by the sum of squared activity across channels is an effective way to reduce the statistical dependence between channels[23,24]. Thus, this type of normalization will produce a stimulus representation that is very efficient, a valuable feature for a versatile perceptual visual system that must resolve different tasks such as relative motion estimation, identification or recognition of the same object.

Our findings strongly suggest that the main difference between perception and action is a different form of the divisive normalization stage that controls the sensitivity of each system. Our modeling effort demonstrates that the different behaviors can be achieved by changing only the normalization pool of this nonlinear stage. For perception, a large normalization pool has been proposed for push-pull mechanisms observed with center-surround interactions[23,24] or inhibition between similar or different orientation or direction channels[24]. It has also been proposed for masking interactions between different orientation channels in humans[34] and appears to be a generic mechanism for controlling the integration of different signals[35,36] (see also ref. 10). In particular, divisive normalization can be adjusted to account for either integration by weighted pooling or segmentation, through winner take all, when similar or too largely different information is presented[35]. Such a shift between integration and segmentation has been also observed for ocular following[32] (see ref. 17 for a review) and motion perception[33]. The difference that we observed between perception and action suggests that the brain could extract the task-relevant information through divisive normalization. Thus, rather than relying on different mechanisms, perceptual and motion systems could simply integrate information differently through an adjustable, task-dependent tuning of the normalization pool.

In our model, gain control is performed across spatiotemporal filters; that is, before the decoding of speed information. This is consistent with previous computational and physiological studies[24,35] that proposed that a first gain control occurs in area V1 to normalize the activity of these neurons before feeding pattern direction and speed-selective cells found in either area V1[28,37] or area MT[38]. Recent studies have linked nonlinear mechanisms, such as gain control, operating at population levels in primary visual cortex with behavioral performance for both perception[39,40] and eye movements[17,41]. In this context, the use of motion clouds transpires to be a powerful approach for elucidating the cascade of linear and nonlinear stages of cortical motion processing in both human and nonhuman primates[13].

Finally, it should be noted that our results are, to the best of our knowledge, the first direct evidence that richer stimuli, similar to natural images, trigger quicker, stronger and more reliable motor responses in humans. Over the last decade, a very large set of studies has supported the view that early neural population activities are sparser and more reliable when using natural images[38]. Moving textures, such as those introduced here, could be very useful to further investigate neural population responses in and across speeds and explain why most of monkeys' speed-sensitive units exhibit more reliable and narrow tuning when presented with random dots instead of simple gratings[28]. The behavioral consequences of such optimal processing are, however, still largely unknown. We found that simple oculomotor responses can be used to probe the consequences of manipulating the statistical properties of motion inputs, from artificial grating-like to more natural-like textures. Using a parametric approach, we observed that the open-loop phase of tracking eye movements becomes stronger and more reliable across trials. This result questions how sparse, less-variable population responses can be transformed into faster and stronger motor responses. Our model did not attempt to address this specific question, instead focusing on response sensitivity rather than absolute amplitude of eye movements. Future work will be needed to investigate this aspect, strengthening the importance of ocular following for probing the details of visuomotor transformations[17].

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J. Colombet, F. Barthélemy and X. DeGiovanni for their excellent technical support and A. Meso for improving the readability of the manuscript. We are grateful to Y. Frégnac, P. Cavanagh, K. Gegenfurtner, A. Movshon and T. Freeman for helpful comments and discussions in the preparation of the manuscript. This work was supported by the EU grant CODDE (VIIth Framework, Marie Curie Program, PITN-2008-214728), by the Centre National de la Recherche Scientifique and by the EU projects FACETS (VIth Framework, IST-FET-2005-15879) and BrainScales (VIth Framework, IST-FET-2011-269921).

AUTHOR CONTRIBUTIONS

G.S.M. and P.M. directed the study. All of the authors conceived the experiments. L.U.P. designed the motion texture stimuli. C.S. and A.M. performed the experiments and data analysis. P.M. and L.U.P. designed the model. G.S.M. and P.M. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nn.3229.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Milner, A.D. & Goodale, M.A. *The Visual Brain in action* 2nd edn (Oxford University Press, 2006).
2. Cardoso-Leite, P. & Gorea, A. On the perceptual/motor dissociation: a review of concepts, theory, experimental paradigms and data interpretations. *Seeing Perceiving* **23**, 89–151 (2010).
3. Spering, M. & Montagnini, A. Do we track what we see? Common versus independent processing for motion perception and smooth pursuit eye movements: a review. *Vision Res.* **51**, 836–852 (2011).
4. Beutter, B.R. & Stone, L.S. Human motion perception and smooth eye movements show similar directional biases for elongated apertures. *Vision Res.* **38**, 1273–1286 (1998).
5. Krauzlis, R.J. & Adler, S.A. Effects of directional expectations on motion perception and pursuit eye movements. *Vis. Neurosci.* **18**, 365–376 (2001).
6. Churchland, A.K. *et al.* Directional anisotropies reveal a functional segregation of visual motion processing for perception and action. *Neuron* **37**, 1001–1011 (2003).
7. Spering, M., Pomplum, M. & Carrasco, M. Tracking without perceiving: a dissociation between eye movements and motion perception. *Psychol. Sci.* **22**, 216–225 (2011).
8. Tavassoli, A. & Ringach, D.L. When your eyes see more than you do. *Curr. Biol.* **20**, R93–R94 (2010).
9. Spering, M. & Gegenfurtner, K.R. Contrast and assimilation in motion perception and smooth pursuit eye movements. *J. Neurophysiol.* **98**, 1355–1363 (2007).
10. Ohshiro, T., Angelaki, D.E. & DeAngelis, G.C. A normalization model of multisensory integration. *Nat. Neurosci.* **14**, 775–782 (2011).

11. Carandini, M. & Heeger, D.J. Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**, 51–62 (2012).
12. Schrater, P.R., Knill, D.C. & Simoncelli, E.P. Mechanism of visual motion detection. *Nat. Neurosci.* **3**, 64–68 (2000).
13. Léon, P.S., Vanzetta, I., Masson, G.S. & Perrinet, L.U. Motion clouds: model-based stimulus synthesis of natural-like random textures for the study of motion perception. *J. Neurophysiol.* **107**, 3217–3226 (2012).
14. Miles, F.A., Kawano, K. & Optican, L.M. Short-latency ocular following responses of monkey I. Dependence on temporospatial properties of visual input. *J. Neurophysiol.* **56**, 1321–1354 (1986).
15. Gellman, R.S., Carl, J.R. & Miles, F.A. Short-latency ocular following responses in man. *Vis. Neurosci.* **5**, 107–122 (1990).
16. Masson, G.S. & Castet, E. Parallel motion processing for the initiation of short-latency ocular following in humans. *J. Neurosci.* **22**, 5149–5163 (2002).
17. Masson, G.S. & Perrinet, L.U. The behavioral receptive field underlying motion integration for primate tracking eye movements. *Neurosci. Biobehav. Rev.* **36**, 1–25 (2012).
18. Britten, K.H. *et al.* Responses of neurons in macaque MT to stochastic motion signal. *Vis. Neurosci.* **10**, 1157–1169 (1993).
19. Lisberger, S.G.L. Visual guidance of smooth-pursuit eye movements: sensation, action and what happens in between. *Neuron* **66**, 477–491 (2010).
20. Green, D.M. & Swets, J.A. *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
21. McKee, S.P. A local mechanism for differential velocity detection. *Vision Res.* **21**, 491–500 (1981).
22. Reisbeck, T.E. & Gegenfurtner, K.R. Velocity tuned mechanisms in human motion processing. *Vision Res.* **39**, 3267–3285 (1999).
23. Hautus, M.J. & Meng, X. Decision strategies in the ABX (matching-to-sample) psychophysical task. *Percept. Psychophys.* **64**, 89–106 (2002).
24. Heeger, D.J. Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* **9**, 181–197 (1992).
25. Schwartz, O. & Simoncelli, E.P. Natural signal statistics and sensory gain control. *Nat. Neurosci.* **4**, 819–825 (2001).
26. Barthélemy, F.V. *et al.* Dynamics of distributed 1D and 2D motion representations for short-latency ocular following. *Vision Res.* **48**, 501–522 (2008).
27. Perrone, J.A. & Thiele, A. Speed skills: measuring the visual speed analyzing properties of primate MT neurons. *Nat. Neurosci.* **4**, 526–532 (2001).
28. Priebe, N.J., Lisberger, S.G. & Movshon, J.A. Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *J. Neurosci.* **26**, 2941–2950 (2006).
29. Jazayeri, M. & Movshon, J.A. Optimal representation of sensory information by neural populations. *Nat. Neurosci.* **9**, 690–696 (2006).
30. Jazayeri, M. & Movshon, J.A. Integration of sensory evidence in motion discrimination. *J. Vis.* **7**, 1–7 (2007).
31. Schenk, T., Franz, V. & Bruno, N. Vision-for-perception and vision-for-action: which model is compatible with the available psychophysical and neuropsychological data? *Vision Res.* **51**, 812–818 (2011).
32. Sheliga, B.M., Kodaka, Y., FitzGibbon, E.J. & Miles, F.A. Human ocular following initiated by competing image motions: evidence for a winner-take-all mechanism. *Vision Res.* **46**, 2041–2060 (2006).
33. Webb, B.S., Ledgeway, T. & McGraw, P.V. Relating spatial and temporal orientation pooling to population decoding in human vision. *Vision Res.* **50**, 2274–2283 (2010).
34. Candy, T.R., Skoczenski, A.M. & Norcia, A.M. Normalization models applied to orientation masking in the human infant. *J. Neurosci.* **21**, 4530–4541 (2001).
35. Rust, N.C., Mante, V., Simoncelli, E.P. & Movhson, J.A. How MT cells analyze the motion of visual patterns. *Nat. Neurosci.* **9**, 1421–1431 (2006).
36. Busse, L., Wade, A.R. & Carandini, M. Representation of concurrent stimuli by population activity in visual cortex. *Neuron* **64**, 931–942 (2009).
37. Rust, N.C., Schwartz, O., Movshon, J.A. & Simoncelli, E.P. Spatiotemporal elements of macaque V1 receptive fields. *Neuron* **46**, 945–956 (2005).
38. Vinje, W.E. & Gallant, J.L. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* **287**, 1273–1276 (2000).
39. Tsai, J.J., Wade, A.R. & Norcia, A.M. Dynamics of normalization underlying masking in human visual cortex. *J. Neurosci.* **32**, 2783–2789 (2012).
40. Chen, Y., Geisler, W.S. & Seidemann, E. Optimal decoding of correlated neural population responses in the primate visual cortex. *Nat. Neurosci.* **9**, 1412–1420 (2006).
41. Reynaud, A., Masson, G.S. & Chavane, F. Dynamics of local input normalization result from balanced short and long-range intra-cortical interactions in area V1. *J. Neurosci.* **32**, 12558–12569 (2012).

ONLINE METHODS

Experiments were performed on four naive participants (three males, one female) and one of the authors (C.S.) who gave their informed consent. Participants had normal visual acuity. The experiments were conducted in compliance with the declaration of Helsinki and approved by the INT Ethical Committee.

Stimuli. Moving texture stimuli are a class of band pass–filtered white-noise stimuli[11]. In Fourier space, the envelope of the filter is a Gaussian in the coordinates of the relevant perceptual axis: speed, frequency and orientation[13]. It is fully characterized on these respective axes by its mean and bandwidth. The latter is defined as the s.d. and is equal to the full bandwidth at half-height divided by ~2.35. Thus, a given image (I) is defined by the following equation (see ref. 13 for more details):

$$I = F^{-1}\left[\exp\left(-\frac{1}{2}\left(\frac{f_x.V_x + f_y.V_y + f_t}{B_v.f_r}\right)^2\right)\cdot\frac{1}{f_r}\exp\left(-\frac{1}{2}\left(\frac{\ln\left(\frac{f_r}{sf_0}\right)}{\ln\left(\frac{sf_0 + B_{sf}}{sf_0}\right)}\right)^2\right)\cdot\exp(i\Phi)\right] \quad (1)$$

where F is the Fourier transform, $v = (v_x, v_y)$ the central motion, $f_r = \sqrt{f_x^2 + f_y^2 + f_t^2}$ the radial frequency and Φ is a uniformly distributed phase spectrum in $[0, 2\pi)$.

Central spatial (sf_0) and temporal (tf_0) frequencies were set to define a preferred speed $v = tf_0/sf_0$. We selected the volume spanned by tilting the speed plane with bandwidth B_v. In all experiments, parameter B_v was kept constant to 5% relative to the central velocity v. For ocular following, we used four different speeds (20–80° s^{-1}). For speed discrimination, the reference speed was set to 20° s^{-1}. Spatial frequency bandwidth (B_{sf}) around the central spatial frequency (sf_0) is defined by its projection on the spatial frequency axis (B_{sf}) so that bandwidths can be compared across textures moving at different speeds (**Fig. 1a**). We defined the Gaussian envelope on a logarithmic frequency scale and used six different B_{sf} (0.025, 0.05, 0.1, 0.2, 0.4 and 0.8 cpd) for four different spatial frequencies (0.1, 0.2, 0.4 and 0.8 cpd). Third, all orientations were equally selected, yielding a toroidal envelope. Finally, the envelope was used to linearly filter a white-noise stimulus drawn from a uniform distribution. The stimuli were built with the Scipy library, calibrated and displayed with Psychotoolbox v3 (ref. 42) for Matlab on a CRT monitor (1,280 × 1,024 pixels) driven by a Macintosh MacPro at a refresh rate of 100 Hz. Stimuli covered 27° of visual angle at a viewing distance of 57 cm. Two example movies are shown (**Supplementary Movies 1** and **2**) for small and large bandwidths, respectively.

Eye movements. Right eye position was recorded at 1 kHz using the EYELINK 1000 video eye tracker. Participants' heads were maintained firmly by chin and forehead rests. Eye position signal was calibrated at the beginning of each experimental session. The experiment and data collection were controlled with the Eyelink Toolbox[43]. The behavioral procedure (**Fig. 1a**) is described elsewhere[14–17]. For a given speed or spatial frequency, motion directions (right and leftward) and spatial frequency bandwidths were fully randomly interleaved. On average, 150 trials were collected per condition for each participant.

Speed discrimination. A constant stimuli method (**Fig. 1b**) with stimulus speed reference (S1) at 20° s^{-1} and stimulus speed target (S2) at 20 ± 1, 2, 4, 8 or 16° s^{-1} was used for the behavioral task. Participants were required to fixate the center of the screen. A fixation point was displayed in this location for 200 ms. If the fixation point's color was black, the following was S1; if it was red, the stimulus to be presented was S2. The presentation's order randomly changed on every trial. When the fixation point disappeared, the motion stimulus was displayed for 250 ms, after which a gray screen was presented, ending the trial. S1 and S2 had the same physical characteristics (that is, same sf_0 and B_{sf}) but differed in speed. The participants' task was to determine whether S2 was slower or faster than S1 in a two-alternative forced choice to be selected by clicking the mouse. In a control task, the reference stimulus was a standard grating of same mean spatial frequency (sf_0) as the moving texture and was always presented in the second temporal window.

Matching-to-sample (ABX) experiment. To determine whether participants based their judgments on speed information rather than temporal frequency, we carried out a pattern discrimination task using the ABX task[23]. Participants were presented with two known random texture movies (A,B) and an unknown sample (X). For a given block of trials, stimulus A remained fixed at one mean spatiotemporal frequency and bandwidth, whereas stimulus B was varied in the spatiotemporal frequency space along either the diagonal of constant speed (condition 1) or a line orthogonal to it (condition 2). Note that the latter condition resulted in a change of the mean speed of stimulus B relative to stimulus A. The two conditions were run alternately, in random order. The same range of spatiotemporal frequencies (stimulus B) was tested in the two conditions. All stimuli were displayed with the same parameters as used in the original speed discrimination tasks. Stimulus duration was 250 ms and the mean speed was 20° s^{-1}.

Data analysis. Eye-position data were low-pass filtered using a 6-poles Butterworth filter before eye velocity was computed. Trials containing small saccadic eye movements during the open-loop period (<160 ms after stimulus onset) were discarded. For quantitative analysis, mean eye velocity during a 100–130-ms time window was computed for each trial together with its distribution across trials to estimate the variability of ocular responses. For speed discrimination, 500 trials were collected for each condition and each participant. Psychometric curves were fitted with a cumulative Gaussian function using the Psignifit toolbox[44]. Contrast response functions, linking response amplitude (ocular following) or sensitivity (psychophysics) with stimulus contrast were fitted with the Naka-Rushton equation[16,17,32,45]

$$R = \frac{c^n}{c^n + c_{50}^n} \quad (2)$$

where R is the normalized response (between 0 and 1), c is contrast, n is the slope of the function and c_{50} is its inflexion point (also known as the half-saturation contrast). All fits were significant ($\chi^2 < 0.009$, degrees of freedom = 2, $P < 0.05$).

Model description. A simple model can help understand the difference between oculomotor and perceptual performance. This model is composed of three stages: encoding of the stimulus, gain control and decoding of speed on the basis of maximum-likelihood estimation. Apart from the gain-control stage, the models are identical for both oculomotor and perceptual tasks.

The encoding stage of the model consists of representing the local motion energy in a bank of spatiotemporal channels. Because our stimuli always move along the same direction, we restricted space to a single dimension that we labeled x and its Fourier transform f_x. When both spatial and temporal frequencies are represented on logarithmic scales, stimuli that have identical speeds lie on a line with a slope of 1 in this spatiotemporal log-frequency space. A channel sensitive to speed will therefore be an oriented filter in this space (for example, see ref. 46). We modeled this oriented filter as a bivariate normal distribution; that is, for a given channel φ_i,

$$\varphi_i(f_x, f_t) \sim N(\mu, \Sigma) \quad (3)$$

where

$$\mu = \begin{pmatrix} f_{x,i} \\ f_{t,i} \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_t \\ \rho\sigma_x\sigma_t & \sigma_t^2 \end{pmatrix} \quad (4)$$

Each channel is fully determined by its location (center spatial $f_{x,i}$ and temporal $f_{t,i}$ frequencies), spread (variance in spatial σ_x^2 and temporal σ_t^2 frequencies) and a correlation coefficient ρ that determines the concentration along the diagonal line of constant velocity (when it is 0, the filter is separable in spatial and temporal frequencies; when it is 1, the filter is perfectly selective for only one speed for a range of spatial and temporal frequencies). For simplicity, we assumed that our channels ($N = 441$) are homogeneously distributed in spatiotemporal log-frequency space[22,27,28]. In our implementation, the channels uniformly tile the spatiotemporal log-frequency space, with spacing between two consecutive center spatial frequencies Δf_x and between two consecutive center temporal frequencies Δf_t (in our implementation of the model, we arbitrarily set Δf_x and Δf_t to 0.1 $\log_2$ units, thus spanning 21 different speeds from 4 to 64° s^{-1}; **Fig. 6a**). To have the same number of spatiotemporal channels for each speed, we chose the tiling

to occupy a diamond-shaped domain in spatial and temporal frequency space (**Fig. 6b**). The specific shape of the tiling ensured that the sum over all speed channels was a constant over speed, a pre-requisite to use the likelihood decoding scheme introduced below[47]. We also assume that the spreads σ_x^2 and σ_t^2 are constant in log-frequency space, an assumption consistent with psychophysical and neurophysiological data[22,48]. We leave these spread values as free parameters of our model, and, to reduce the number of free parameters, we assume that the spreads are identical across spatial and temporal dimensions, and thus label this common spread σ^2.

When contrast increases, the speed tuning of neurons in area MT gets sharper[28]. We implemented this feature by increasing the value of the correlation coefficient ρ in the bivariate normal that represented each channel. A glance at our data reveals that small variations at low contrast have more effects on eye movements and perception than relatively large variations at high contrast, suggesting some kind of compressive nonlinearity between stimulus contrast and channel correlation coefficient. We implemented this nonlinearity as follows:

$$\rho = \left(1-(1-c)^{\gamma}\right)^{1/\gamma} \tag{5}$$

where c is the stimulus contrast and γ is a free parameter of the model characterizing the strength of the nonlinearity. The output of a spatiotemporal channel φ_i to a stimulus s with contrast c and bandwidth b is

$$m_i(c,b) = \iint \varphi_i(f_x, f_t) * s(c,b)\, df_x df_t \tag{6}$$

The second stage of the model consists in a gain control of the output of the spatiotemporal channels. Such a gain adjustment is typically done by a gain control scheme known as divisive normalization[23,24] and has been critically involved in many context-dependent sensory integration processes[10]. The only difference in the oculomotor and perception models lies in the nature of the suppressive field for the gain control computation. We used a Naka-Rushton[45] type of gain control. For eye movements, the suppressive field was taken to be identical to the excitatory field

$$n_i(c,b) = \frac{m_i^2(c,b)}{m_e^2 + m_i^2(c,b)} \tag{7}$$

where m_e is the semi-saturation constant for the eye-movement task. This simple computation simply avoids out-of-bounds activities. In contrast, for perception, the suppressive field is taken to be the sum of the activity of all channels

$$n_i(c,b) = \frac{m_i^2(c,b)}{m_p^2 + \sum_j m_j^2(c,b)} \tag{8}$$

where m_p is the semi-saturation constant for the perceptual task. When the stimulus bandwidth increases, it will activate relatively more channels and, as a result, the suppressive field will become larger. In other words, as the stimulus bandwidth increases, the evidence for each speed diminishes.

Although the gain control is a simple local computation in the eye-movement task, the computation takes into account the global activity of the network in the perceptual task. One potential biologically plausible reason for this dissociation is the incentive to implement a fast computation in the eye-movement task. In contrast, the perceptual task is less time critical and it is arguably more important to preserve the whole richness of the stimulus, particularly by combining the information across all spatiotemporal channels.

The third stage of the model consists of decoding the speed information and it is again common to both tasks. We choose here to estimate the likelihood of the speed distribution, similarly to what has been done previously to decode perceived motion direction[29,30]. Note that other decoding schemes such as vector averaging[19] could be equally effective given that our stimuli are symmetric about their mode[33]. The logarithm speed likelihood of each spatiotemporal channel was computed as the product of the response of the stimulus in a given channel and the logarithm of the speed tuning function of that channel

$$\log p(n_i \mid s) = n_i(c,b)\log(\psi_i) \tag{9}$$

where Ψ_i is the speed tuning of channel i centered on speed $v_i = f_{t,i}/f_{x,i}$. Because the channels are bivariate normal distributions, the speed tuning function of a channel is its cross-section along the $f_t = -f_x$ axis

$$\psi_i(v) = \exp\left(-(v-v_i)^2 \Big/ \left(2\sigma^2(1-\rho)\right)\right) \tag{10}$$

Log-likelihoods are then summed across all channels to give the overall speed log-likelihood of the stimulus

$$\log(L(s)) = \sum_{i=1}^{N} n_i(c,b)\log(\psi_i) \tag{11}$$

For the ocular following task, the overall speed likelihood offers an estimate of the speed sent to the oculomotor plant (the mode of the distribution) and of the reliability of this estimate (the spread of the distribution). Thus, the inverse of the s.d. of the speed likelihood is a good estimate of the sensitivity of the ocular following. Consistent with our behavioral results, ocular sensitivity in our model increases with stimulus bandwidth. The reason is that the overall speed likelihood gets more concentrated around its mode as more spatiotemporal channels are activated when the stimulus bandwidth increases. In addition, ocular sensitivity also increases when contrast increases, simply because more energy is presented.

For the perception task, the overall speed likelihood provides a distribution of the relative evidence for each speed in the stimulus. When two stimuli of slightly different speeds are compared, the task of the visual system resumes comparing the degree of overlap between the two speed likelihoods. Intuitively, the more the two likelihoods overlap, the more difficult the discrimination between the two speeds. More formally, we take as the measure of similarity between the two likelihoods the area under the receiver operating characteristic (ROC) curve[20]. This ROC is obtained by taking an arbitrary criterion, computing the probabilities that stimuli S1 and S2 exceed this criterion (this is one point on the ROC curve), and repeating the procedure for other criteria. The probability $p(s_2 > s_1)$ that stimulus with center speed v_2 is perceived faster than a stimulus with center speed v_1 is given by the area under the ROC curve. Repeating this procedure for multiple test speeds s_2 for a given standard speed s_1, we built a modeled psychometric function and fitted this function with a cumulative Gaussian distribution. The inverse of the spread of the best-fit Gaussian determines the sensitivity of the model for the psychophysical task. Consistent with our behavioral results, perceptual sensitivity in our model does decrease with stimulus bandwidth. The reason is that the overall speed likelihood gets weaker evidence from a larger number of spatiotemporal channels as stimulus bandwidth increases. In addition, perceptual sensitivity does increase when contrast increases, again simply because more energy is presented.

Note that we have not modeled the biases in tracking speeds that are systematically slower than the speed of the stimulus. There might be several sources for these biases, but they might be modeled if we had included a prior for slow speeds in our model[49]. The prior will introduce stronger biases when the likelihood is more broadly tuned; that is for small stimulus bandwidth and small stimulus contrast. This is similar to what was observed behaviorally. Alternatively, the bias for slow speed might be the result of a non-uniform representation of speed channels[50].

42. Pelli, D.G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
43. Cornelissen, F.W., Peters, E.M. & Palmer, J. The eyelink toolbox: eye tracking with MATLAB and the psychophysics toolbox. *Behav. Res. Methods Instrum. Comput.* **34**, 613–617 (2002).
44. Wichmann, F.A. & Hill, N.J. The psychometric function. I. Fitting, sampling and goodness-of-fit. *Percept. Psychophys.* **63**, 1293–1313 (2001).
45. Sclar, G., Maunsell, J.H. & Lennie, P. Coding of image contrast in central visual pathways of the macaque monkey. *Vision Res.* **30**, 1–10 (1990).
46. Anderson, S.J. & Burr, D.C. Spatial and temporal selectivity of the human motion detection system. *Vision Res.* **25**, 1147–1154 (1985).
47. Dayan, P. & Abbott, L.F. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, 2001).
48. Priebe, N.J., Cassanello, C.R. & Lisberger, S.G. The neural representation of speed in macaque Area MT/V5. *J. Neurosci.* **23**, 5650–5661 (2003).
49. Weiss, Y., Simoncelli, E.P. & Adelson, E.H. Motion illusions as optimal percepts. *Nat. Neurosci.* **5**, 598–604 (2002).
50. Priebe, N.J. & Lisberger, S.G. Estimating target speed from the population response in visual area MT. *J. Neurosci.* **24**, 1907–1916 (2004).

doi:10.1038/nn.3229