Modeling spatial integration in the ocular following response using a probabilistic framework

Laurent U. Perrinet *, Guillaume S. Masson

Institut de Neurosciences Cognitives de la Méditerranée (INCM-UMR 6193, CNRS) 31, ch. Joseph Aiguier, 13402 Marseille Cedex 20, France

Abstract

The machinery behind the visual perception of motion and the subsequent sensori-motor transformation, such as in ocular following response (OFR), is confronted to uncertainties which are efficiently resolved in the primate's visual system. We may understand this response as an ideal observer in a probabilistic framework by using Bayesian theory [Weiss, Y., Simoncelli, E.P., Adelson, E.H., 2002. Motion illusions as optimal percepts. Nature Neuroscience, 5(6), 598–604, doi:10.1038/nn858] which we previously proved to be successfully adapted to model the OFR for different levels of noise with full field gratings. More recent experiments of OFR have used disk gratings and bipartite stimuli which are optimized to study the dynamics of center–surround integration. We quantified two main characteristics of the spatial integration of motion: (i) a finite optimal stimulus size for driving OFR, surrounded by an antagonistic modulation and (ii) a direction selective suppressive effect of the surround on the contrast gain control of the central stimuli [Barthélemy, F.V., Vanzetta, I., Masson, G.S., 2006. Behavioral receptive field for ocular following in humans: dynamics of spatial summation and center–surround interactions. Journal of Neurophysiology, (95), 3712–3726, doi:10.1152/jn.00112.2006]. Herein, we extended the ideal observer model to simulate the spatial integration of the different local motion cues within a probabilistic representation. We present analytical results which show that the hypothesis of independence of local measures can describe the spatial integration of the motion signal. Within this framework, we successfully accounted for the contrast gain control mechanisms observed in the behavioral data for center–surround stimuli. However, another inhibitory mechanism had to be added to account for suppressive effects of the surround.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Visual perception; Motion integration; Tracking eye movements; Ocular following response; Bayesian model; Center–surround interactions

1. Introduction

Local motion signals are often noisy and ambiguous. To elaborate an accurate perception of the spatio-temporal layout of our environment or to direct our actions towards a particular object of interest, the biological system responsible for the visual perception of motion must integrate these piecewise measurements in a selective way. The machinery behind this dynamical integration process has been scrutinized at different levels, from single neurons to behavior. However, one still lacks an integrative view of motion integration.

Probabilistic representations of motion are useful tools to understand these mechanisms. In these representations, all the information extracted from the image is coded as probability distribution functions (PDFs) of the different possible velocities of translation. As is inspired by the architecture of visual cortical area MT, spatio-temporal distributions of activity within large populations of units broadly tuned for direction selectivity build maps of the local motion information (Nowlan and Sejnowski, 1995), these distributed representations may be interpreted as maps of PDFs. An ideal decision process could then be used to infer the most plausible interpretation of the image from this distributed probabilistic representation using

---

* Corresponding author.
  E-mail address: Laurent.Perrinet@incm.cnrs-mrs.fr (L.U. Perrinet).

standard operations from probability theory in order to trigger and control adaptive behavioral responses.

We use such an approach to model how motion information is quickly integrated in order to drive reflexive eye movements involved in tracking object motion in primates. Because the aim of these responses is to stabilize onto the retinas the images of a single object of interest, they are an ideal candidate to probe both the low-level motion mechanisms that single-out this object and compute its motion and the decision process that triggers a motor response in the appropriate direction and at the correct speed.

2. Ocular following responses in primates: a probe of dynamical motion integration

In human and non-human primates, a brief (<200 ms) translation of the visual scene elicits reflexive tracking eye movements at ultra-short latency (~55 ms in monkeys and ~85 ms in humans, see Fig. 1a) (Miles et al., 1986; Gellman et al., 1990). The responses, also called ocular following responses (OFR), exhibit many of the properties that are attributed to low-level motion detectors. They are triggered primarily by a motion energy signal that is
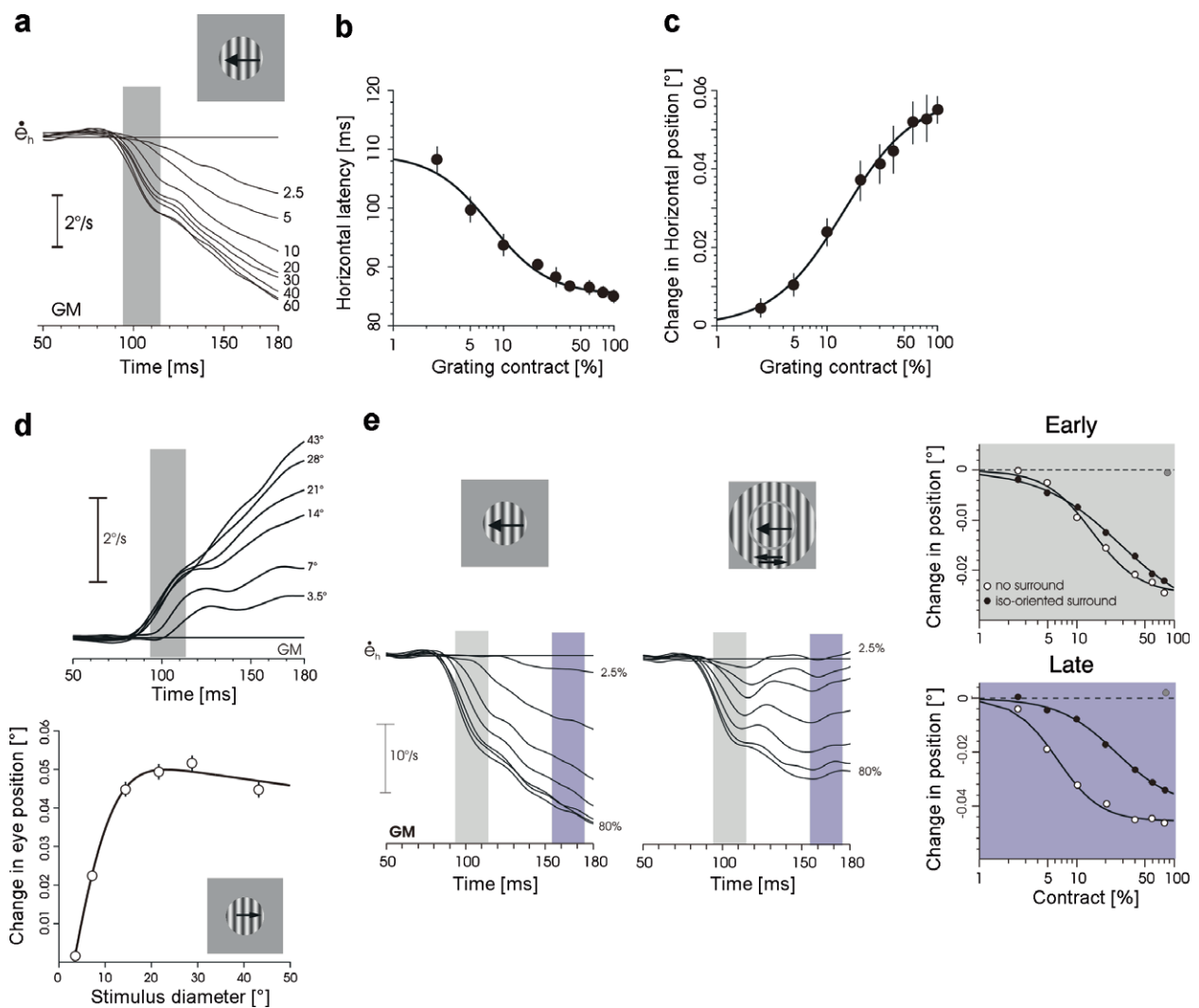

Fig. 1. Basic properties of human OFR. Several properties of motion integration for driving ocular following as summarized from our previous work. (a) A leftward drifting grating elicits a brief acceleration of the eye in the leftward direction. Mean eye velocity profiles illustrate that both response amplitude and latency are affected by the contrast of the sine-wave grating, given by numbers at the right-end of the curves. Quantitative estimates of the sensori-motor transformation are given by measuring the response amplitude (i.e. change in eye position) over a fixed time window, at response onset. Relationships between (b) response latency or (c) initial amplitude and contrast are illustrated for the same grating motion condition. These curves define the contrast response function (CRF) of the sensori-motor transformation and are best fitted by a Naka–Rushton function (reprinted from (Barthélemy et al., 2007)). (d) At fixed contrast, the size of the circular aperture can be varied to probe the spatial summation of OFR. Clearly, response amplitude first linearly grows up with stimulus size before reaching an optimal size, the integration zone. For larger stimulus sizes, response amplitudes are lowered (reprinted from (Barthélemy et al., 2006)). (e) OFR are recorded for center-alone and center–surround stimuli. The contrast of the center stimulus is varied to measure the contrast response function and compute the contrast gain of the sensori-motor transformation at both an early and a late phase during response onset. Open symbols are data obtained for a center-alone stimulus, similar to those illustrated in (c). When adding a flickering surround, ones can see that late (but not early) contrast gain is lowered, as illustrated by a rightward shift of the contrast response function (Barthélemy et al., 2006).

extracted from the spatio-temporal luminance distribution in the image (Miles et al., 1986; Masson et al., 2002; Sheliga et al., 2005). For instance, reversing the contrast of a high density dot-pattern during a one-step apparent motion results in a reversed ocular response, following the direction of the first-order motion signal (Masson et al., 2002). These reflexive responses are best driven by motion signals extracted in the low spatial (<1 cpd) and high temporal (>10 Hz) frequency range. Moreover, latency of ocular following depends on the temporal frequency of the local change in luminance and not the speed of the visual motion (Miles et al., 1986; Gellman et al., 1990).

Contrast is another important factor for driving ocular following responses, as it is related to the signal-to-noise ratio of the mechanism that computes a velocity signal with appropriate strength and accuracy to drive the sensori-motor transformation. Both latency and amplitude of the earliest phase of ocular following were found to be nonlinearly dependent upon the contrast of moving sine waves (Fig. 1b,c). When considering the amplitude of the initial eye acceleration, the contrast response functions mimics those observed for motion selective neurons: a rapidly expanding phase at low contrast followed by a saturation of response amplitude with high contrast (Masson and Castet, 2002; Sheliga et al., 2005; Barthélemy et al., 2007). This relationship is best fitted by a Naka–Rushton function, similar to that used for describing contrast response functions of neurons at various stages along the motion pathway (Sclar et al., 1990; Albrecht et al., 2002). A similar relationship was found between response amplitude and signal-to-noise ratio when varying the percentage of correlated motion in a dynamic sequence of random-dot patterns (Masson, Barthélemy and Vanzetta, unpublished). Latency of ocular following also undergoes a considerable change when contrast varies from low (<5%) to mid-range values. Barthélemy et al. (2007) showed in humans that the relationships between response latency and contrast can be best described by an inverted Naka–Rushton function.[1] Lastly, higher order motion cues can influence the direction of tracking initiation, albeit with a slightly longer latency, reflecting the temporal dynamics of motion integration in the primate visual motion pathway (Masson and Castet, 2002; Masson, 2004).

These short-latency ocular following responses are driven by a global motion signal built by pooling local motion over a very large part of the visual field (Masson, 2004; Barthélemy et al., 2006). Masson and colleagues have investigated the properties of such spatial integration of motion in both humans (Barthélemy et al., 2006) and monkeys (Barthélemy and Masson, 2006). They found that motion is linearly integrated over the central 20° of the visual field. Within this central, driving part of the visual field, different motion signals are linearly integrated by

computing a vector average of the local motion directions (Masson and Castet, 2002; Barthélemy et al., 2006). However, at high spatial frequency, stretching the stimulus size above the optimal summation area results in a saturation followed by a decrease of initial eye velocity. Reductions in eye velocity with large stimuli were much stronger in monkeys than in humans. Moreover, hyper-saturation was seen mostly for long temporal integration window, reflecting the fact that surround suppression is delayed relative to the center-driven response onset.

Such an hyper-saturation is generally interpreted as the signature of an inhibitory surround, driven by motion in the same direction as the center. Such surround suppression was originally described by Miles and colleagues in monkeys (Miles et al., 1986). Using very large random-dots patterns, they found that drifting the surround (typically larger than 40° of visual angle) in the same direction lowered the responses driven by central motion, a phenomenon called "in-phase suppression". On the contrary, antagonistic surround motion boosted the initial eye acceleration: "anti-phase enhancement". Both suppression and enhancement were evident only in the later part of the responses, i.e. ~30 ms after response onset. Similar, albeit much smaller effects were reported in humans (Gellman et al., 1990). In monkeys, similar modulatory effects of surround motion has been demonstrated at the level of both areas MT (Born and Tootell, 1991) and MST (Eifuku and Wurtz, 1999).

In humans, Barthélemy et al. (2006) found further evidence for such surround inhibition when comparing contrast response functions obtained with or without a dynamical surround at high contrast. They used a flickering pattern as surround stimulus to disentangle visual and motor interactions in center–surround conditions. A flickering sine-wave is a static grating whose contrast is time modulated and which can be interpreted as the sum of two similar gratings, drifting in opposite directions. Thus, no net motion signal was present in the surround area, which therefore did not elicit ocular responses when presented alone. Nevertheless, surround flicker lowered contrast gain for center-driven responses, resulting in flatter and shifted amplitude-contrast response functions (see Fig. 1e). We have obtained similar results in macaque ocular following responses. Interestingly, this inhibitory effects of surround motions was delayed relative to tracking onset and built up over time in both humans and monkeys. Overall, these inhibitory surround effects are very similar to that observed at the single neuron level (Cavanaugh et al., 2002) or at population level (Reynaud et al., 2007) in primate area V1. Similar modulatory surround have been reported in area MT (Heuer and Britten, 2002) but no data are yet available for area MST in macaques.

In monkeys, ocular following responses are initiated shortly after the earliest response onset of neurons in the medio-temporal (MT) and medial superior (MST) areas of the dorsal pathway. Kawano (1999) and colleagues have tracked the temporal dynamics of neuronal information

---

[1] This suggests a generic monotonous relationship between contrast – that is signal-to-noise ratio – and time of integration in this early phase of information integration.

flow from cortical to sub-cortical structures, leading to eye movement onset. Furthermore, complete lesions of areas MT and MST in the macaque dorsal pathway completely abolish ocular following as well as other short-latency tracking responses such as disparity- and radial flow vergence (Takemura et al., 2007). Firing rates of neurons in both areas MT and MST encode for stimulus direction and speed when presented with the stimuli driving ocular following (Kawano et al., 1994). Discharge patterns of MST neurons reflect the different components of the visual signal such as image acceleration and velocity. On the other hand, properties of ocular responses (i.e. eye velocity profiles) are best correlated with neuronal discharges in the ventral paraflocullus (VPFL) lobe of the cerebellum (Takemura and Kawano, 2002). These simulations suggest that MST neurons represent global dynamic properties of the visual stimulus and therefore that the entire populations of MST neurons encodes visual information for ocular following responses. MST neurons are often seen as template neurons that encode global flow informations by pooling local motion information from MT neurons (Duffy and Wurtz, 1991). A critical step is therefore to understand how local motion information is extracted and pooled at the population level to encode a single vector related to object direction and speed. Similar population decoding scheme have been recently proposed to account of initial eye acceleration for voluntary smooth pursuit eye movements in macaque (Priebe and Lisberger, 2004).

In summary, ocular following responses are an excellent opportunity to probe the dynamics of motion integration at behavioral level. The amplitude of the earliest phase of ocular following is proportional to initial eye acceleration which has been shown to be directly related to image velocity (see (Lisberger et al., 1987)). Moreover, a change in eye speed or direction would reflect a change in the output of the visual motion pathway as evidenced by the parallel dynamics observed for solving the aperture problem at the neuronal (Pack and Born, 2001; Smith et al., 2005) and behavioral (Masson et al., 2000; Masson and Castet, 2002) levels. From the properties of the behavioral responses, we can then infer the basic properties of motion processing and how a single velocity signal is extracted from a cascade of neuronal populations from areas V1, MT and MST. Moreover, since the dynamics of the earliest part of eye movements reflects the firing rate of output neurons of the cortical visual processing (Osborne et al., 2004), we might then illustrate the neuronal dynamics at a fine grain and probe the output of large scale neuronal networks implementing a neuronal solution for motion integration and segmentation.

3. An "ideal observer" model for OFR

We will use here the tools of *statistical inference* to model the behavioral results obtained in both human and non-human primates (Barthélemy and Masson, 2006; Barthélemy et al., 2006, 2007; Masson and Castet, 2002). Our

goal is to model how information is pooled over a large population of broadly tuned neurons to extract a signal related to the retinal velocity of the moving object. As in Barthélemy et al. (2007), we will study the reference signal given by the model (usually the velocity that is considered as the most probable) and compare it to the behavioral data by assuming that the gain of the oculo-motor system (the observed change in eye position) is proportional to this velocity. This assumption is reasonable since we restrict ourselves to ultra-short latencies and that our measurements remain in the open-loop part of the dynamics (inferior than 100 ms in human), that is that due to the latencies in the visual system, there was no feed-back of the eye's motion on the system.

As described in Barthélemy et al. (2007), we model the ideal response $\vec{v}_m$ knowing the given information as the conditional expectancy:

$$\vec{v}_m = E(\vec{v}|I) = \int \vec{v} P(\vec{v}|I) \, dP(\vec{v}|I) \tag{1}$$

where $E(\vec{v}|I)$ is the mean velocity computed from the probability of the different possible velocities $\vec{v}$ knowing the observed image I. A major difficulty is to compute this probability $P(\vec{v}|I)$ globally over the whole field and a solution which seems to be implemented in the visual system consists in pooling the information that is extracted locally and which acts as a distributed probabilistic representation of the map of local translation velocities. For the purpose of generality we define these local measurements as different nodes $\mathbf{n} \in \mathscr{P}$, where $\mathscr{P}$ is the total population of nodes. Typically in the visual system, the nodes have local receptive fields where $P(\vec{v}|I, \mathbf{n})$ may be more easily computed (see Fig. 2). In fact, it is then possible to evaluate locally the probability distribution of the velocity $\vec{v}$ thanks to a stochastic model of the local translation near $\mathbf{n}$: $I_{\mathbf{n}}(\vec{x}, t) = I_{\mathbf{n}}(\vec{x} - \vec{v} \, dt, t - dt) + v$, where $\vec{x}$ is the position in the receptive field and v is a Gaussian noise image of variance $\sigma_{\mathbf{n}}^2$. By definition, this noise is inversely proportional to the Michelson's Contrast and we note the full contrast image: $I_{100,\mathbf{n}} = C^{-1} \cdot I_{\mathbf{n}}$.

Adding a Gaussian prior of variance $\sigma_{p,\mathbf{n}}^2$ favoring slow speed (Weiss et al., 2002), it follows from Bayes' theorem:

$$P(\vec{v}|I, \mathbf{n}) = \frac{1}{Z} \cdot e^{-\frac{C^2 \cdot D \, I_{100,\mathbf{n}}(\vec{v})}{2 \cdot \sigma_{\mathbf{n}}^2}} \cdot e^{-\frac{\|\vec{v}\|^2}{2 \cdot \sigma_{p,\mathbf{n}}^2}} \tag{2}$$

where $DI_{100,\mathbf{n}} = C^{-2} \cdot DI_{\mathbf{n}}$ is the contrast-normalized gradient constraint for the local image in the receptive field:

$$DI_{100,\mathbf{n}}(\vec{v}) = \|I_{100,\mathbf{n}}(\vec{x}, t) - I_{100,\mathbf{n}}(\vec{x} - \vec{v} \cdot dt, t - dt)\|^2 \tag{3}$$

This constraint function is exactly the same as the one used in image processing when computing the hypothesis of conservation of image luminance along different translation velocities. It is equivalent also to the correlation in the spectral space with the plane perpendicular to $\vec{v}$ (Simoncelli and Heeger, 2001). However, one has to preprocess images so that neighboring pixels' luminance values are decorrelated and that such an hypothesis still holds.
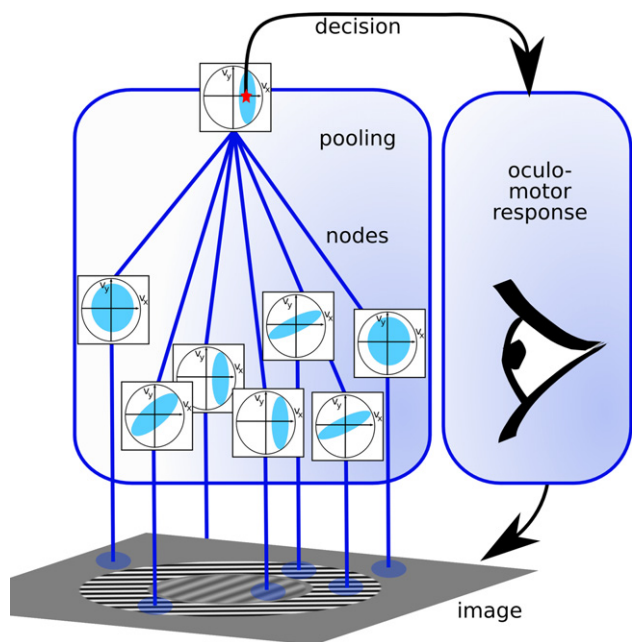
Fig. 2. Architecture of the model. The model consists in pooling elementary cues from local information (represented for instance in neuronal assemblies) to provide a "global" decision which drives the oculo-motor system and closes the oculo-motor loop by moving the eye's position. Every local detector (or *node*) integrates motion information from the image on its receptive field so as to provide a local representation of the possible translation velocity probabilities in the receptive field (schematized by the circular probability distribution functions in velocity space (v_x, v_y)). This distributed information may then pooled across the different nodes (and hence across space) but also integrated in time and potentially by other modalities (e.g. proprioception). A decision (or reference signal) is formed using a statistical inference criterion, usually either the *Maximum A Posteriori* (MAP) probability or the conditional expectancy, and this decision is transformed accordingly by the oculo-motor system to produce an eye movement.

In the range of experiments that we describe here, the whole image will be composed locally by single gratings with parameters their velocity $\vec{v}_\mathbf{n}$ and spatial frequency $f_\mathbf{n}$[2] which may be written:

$$I_{100,\mathbf{n}} = \sin(2\pi f_\mathbf{n}(\vec{x} - \vec{v}_\mathbf{n} \cdot t)) \qquad (4)$$

for which $DI_{100,\mathbf{n}}$ is easy to compute analytically: it is well approached (up to the Nyquist frequency) by a quadratic function with a minimum constraint at $\vec{v} = \vec{v}_\mathbf{n}$. Every node can thus be characterized by a mean $\vec{v}_\mathbf{n}$ and a covariance matrix $\sigma_\mathbf{n}^2$ such that the density $P(\vec{v}|I, \mathbf{n})$ is Gaussian:

$$\mathcal{N}(\vec{v}_\mathbf{n}, \sigma_\mathbf{n}) = \frac{1}{Z} \cdot \exp\left(-\frac{1}{2}(\vec{v} - \vec{v}_\mathbf{n})^T \sigma_\mathbf{n}^{-2}(\vec{v} - \vec{v}_\mathbf{n})\right) \qquad (5)$$

with $Z = \sqrt{\det(2\pi\sigma_\mathbf{n})}$. Qualitatively, information is larger for lower variances.[3] It is useful to determine the principal axis of this distribution so that we have $\sigma_\mathbf{n}^2$ given by

---

[2] To extend this paradigm to natural images, one should consider that this composition is composed locally by multiple gratings as is illustrated in the wavelet representation.

[3] This characterizes the aperture problem as $\sigma_{2,\mathbf{n}} \gg \sigma_{1,\mathbf{n}}$: the variance is higher along the orientation of the line than perpendicular to it.

$$\begin{pmatrix} \cos(\theta_\mathbf{n}) & -\sin(\theta_\mathbf{n}) \\ \sin(\theta_\mathbf{n}) & \cos(\theta_\mathbf{n}) \end{pmatrix} \begin{pmatrix} \sigma_{1,\mathbf{n}}^2 & 0 \\ 0 & \sigma_{2,\mathbf{n}}^2 \end{pmatrix} \qquad (6)$$

where $\theta_\mathbf{n}$ is the orientation of the grating (or the line) relative to the vertical (that is relative to a rightward horizontally perceived motion) and $\sigma_{1,\mathbf{n}} < \sigma_{2,\mathbf{n}}$ are the two eigenvalues of the distribution. In particular, the solution to the gain in Eq. (1) for a full-field grating of speed $\vec{v}_g$ as a function of the contrast, that is the contrast response function (CRF) is thanks to Eq. (2) necessarily a Naka–Rushton curve (Naka and Rushton, 1966) of slope 2:

$$\vec{v}_m(C) = \frac{C^2}{C_{50}^2 + C^2}\vec{v}_g \qquad (7)$$

with $\vec{v}_m$ being the reference speed, perpendicular to the grating and $C_{50} \propto \frac{\sigma_p}{\sigma_{1,\mathbf{n}}}$ is the contrast at half saturation (see Fig. 3, left). This curve gives a general account for the sigmoidal form of the saturation in a wide range of experiments and proved to efficiently account for the OFR at different contrasts compared to the computation with a full field image using Eq. (2) (Barthélemy et al., 2007) (see also Hurlimann et al. (2002) for a similar analysis). It is described by the contrast of half saturation only, however with this quadratic hypothesis it is not possible to fit CRFs which look more or less "binary" (see Fig. 3, right).

We have recently shown that such a Bayesian model can accurately account for a large set of contrast response functions such as obtained with different motion stimuli such as grating, barber-poles (i.e. a grating drifting behind an elongated, tilted aperture) and plaids (i.e. the sum of two gratings moving in different directions). We elaborated a two-pathways version of the inferential model to account for the independent extraction of 1D (i.e. gratings) and 2D (i.e. local features) motion that drive early and late component of ocular following (Masson and Castet, 2002). More important, we found that different contrast gain can be simulated using a shared motion extraction stage across different stimuli. Best fits are shown in Fig. 4a–e. This result shows that the dynamic of ocular following responses reflects how motion information is pooled across different channels extracting grating and feature motions and reflect the signal-to-noise ratio within each channel. Each channel has different temporal integration properties, also fitted to the observed data (Barthélemy et al., 2007). However, the model cannot explain the difference in latency observed between early (i.e. 1D-driven) and late (i.e. 2D-driven) component of ocular following (Masson et al., 2000; Masson and Castet, 2002). This pure delay is another argument for assuming independent nodes in extracting motion signals which are progressively taken in account in the global behavioral signal.

Knowing contrast response functions for a single grating motion (Fig. 4a), we next asked the question of how to combine these local, non-linear responses so as to give a global motion integration. Such a pooling of information
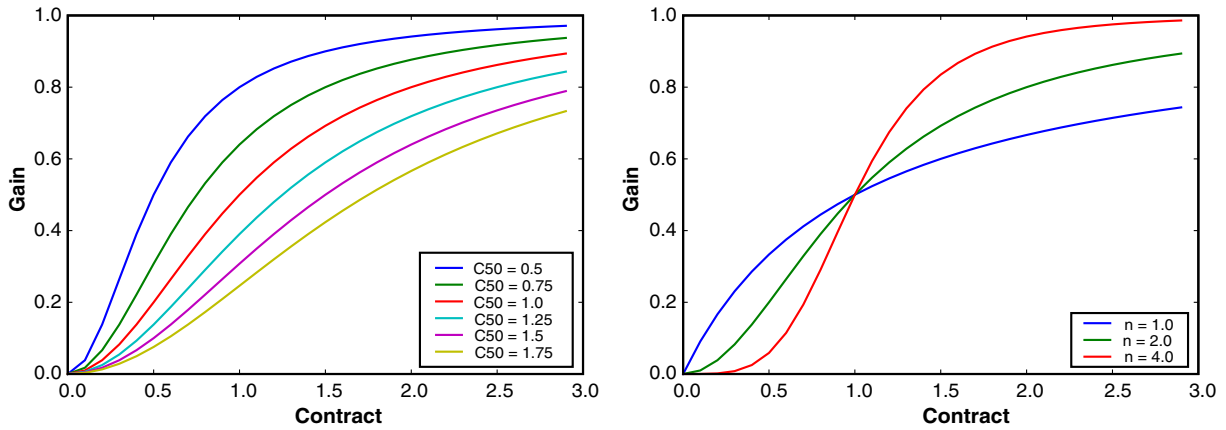
Fig. 3. Naka–Rushton curves. (Left) A typical response to OFR as a function of contrast is shown as a Naka–Rushton curve according to Eq. (7). These curves are totally described by their saturation at half contrast and show a typical sigmoidal shape: the gain is zero at zero contrast, it smoothly increases to reach a linear regime at half saturation contrast and then saturates to one. A property of these curves is that the tangent of the curve at half contrast meets the origin. (Right) Using Naka–Rushton curves with powers of n instead of 2 (here with $C_{50} = 1$), one could produce a richer set of curves which have different slopes at half-contrast, from a "sloppy" response ($n = 1$) to a more "binary" response ($n = 4$).
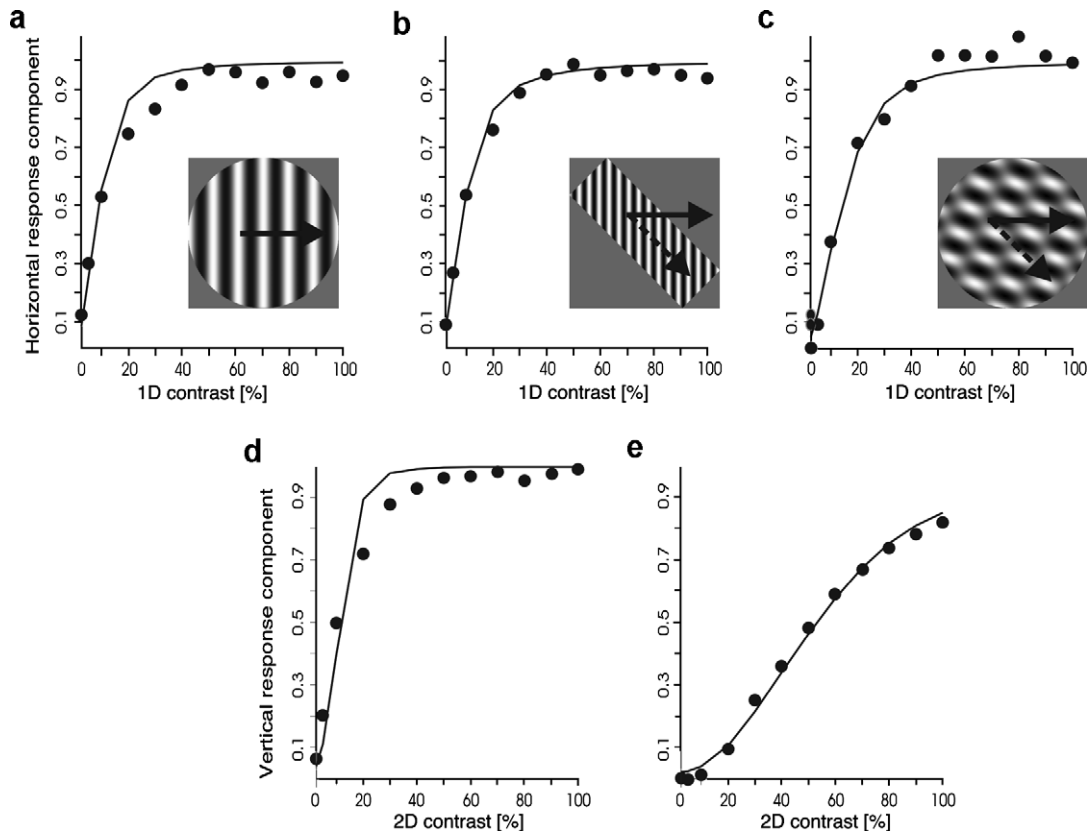

Fig. 4. Contrast response function of OFR. Amplitude of (a, b, c) horizontal and (d, e) vertical ocular following responses plotted against contrast of 1D (i.e. gratings) or 2D (i.e. features such as line-endings or blobs) motion cues. Different contrast gain are observed for each stimulus (i.e. grating, barber-pole and plaid motion) but all relationships can be simulated using a single probabilistic model such as is developed herein. The only modification is the introduction of a second, parallel pathway that extracts 2D feature motions. Re-plotted from Barthélemy et al. (2007).

across space is a key aspect of motion integration for reducing noise or normalizing driving inputs with context information. We have investigated how ocular following depends upon this spatio-temporal integration in two series of experiments. First, we asked what is the optimal stimu-lus size for driving ocular following responses, using grat-ing stimuli of different spatial frequencies. Second, we probed how the contrast response function of OFRs driven by a central moving patch would be affected by surround-ing motion signals. These experiments have been conducted

both in humans (Barthélemy et al., 2006) and monkeys and we will first give a general analytical formulation before applying it to the spatial integration of motion.

4. Integration of independent information for the OFR

A first analytical solution is given thanks to the hypothesis of independence of the measure of velocity on these different nodes (Weiss et al., 2002) which permits to write:

$$\log P(\vec{v}|I) = \sum_{\mathbf{n} \in \mathscr{P}} \log P(\vec{v}|I, \mathbf{n}) \tag{8}$$

This ensures that from Eq. (2), as a product of Gaussian, the resulting distribution is also a Gaussian: $P(\vec{v}|I) = \Pi_{\mathbf{n}} P(\vec{v}|I, \mathbf{n}) = \mathscr{N}_{(\vec{v}_m, \sigma)}(\vec{v})$. In that case, the *Maximum A Posteriori* will again correspond to the mean (see Eq. (1)). Solving the polynomial equation of 2×2 matrices in Eq. (8), it follows that the resulting distribution $\mathscr{N}_{(\vec{v}_m, \sigma)}(\vec{v})$ obeys:

$$\begin{cases} \sigma^{-2} = \sum \sigma_{\mathbf{n}}^{-2} \\ \sigma^{-2} \cdot \vec{v}_m = \sum \sigma_{\mathbf{n}}^{-2} \cdot \vec{v}_{\mathbf{n}} \end{cases} \tag{9}$$

with $\sigma_{\mathbf{n}}^{-2}$ given by

$$\begin{pmatrix} \cos(-\theta_{\mathbf{n}}) & -\sin(-\theta_{\mathbf{n}}) \\ \sin(-\theta_{\mathbf{n}}) & \cos(-\theta_{\mathbf{n}}) \end{pmatrix} \begin{pmatrix} \sigma_{1,\mathbf{n}}^{-2} & 0 \\ 0 & \sigma_{2,\mathbf{n}}^{-2} \end{pmatrix} \tag{10}$$

This simple relation gives a general formula for the pooling of information from different nodes with gaussian PDFs and is a generic tool that we will apply to the different experimental setups. In fact, knowing that a stimulus is constituted by different gratings, it is straightforward using this equation to compute the distribution of the pooled information. However, this will not apply if the stimulus is constituted by transparent overlapping gratings where the probability distribution is multi-modal.

This equation also applies to dynamical integration. In fact, if one assumes independence in time of the measurement noises (the nodes are now centered in space and at different instants in time), the pooling of this information is factorial. In particular, if the probability distribution function is a steady gaussian, Eq. (9) states that the inverse variance will increase proportionally to the integration time (that is from the onset of the integration), so that one may evaluate the dynamical evolution of integration. This was confirmed in a previous study (Barthélemy et al., 2007) and there is thus under this hypothesis a direct equivalence between the integration time and the contrast of the image (see Fig. 1e).

It should be stressed that this is a pure feed-forward model of integration since all information is pooled and no change occur along loops (for instance one does not change the content of the information with intermediate computations as in a recurrent model). It is therefore easy to imagine neuronal implementations, for instance by using divisive normalization (Simoncelli and Heeger, 2001). However, there may be many implementations for the

same function and in particular it is likely that the distributed probabilistic representation is not explicitly coded in the activity of neuronal assemblies. Nevertheless, our general model has more predictive power from our knowledge of the representation and it's link to statistical inference and Bayesian theory in general. In general, from the commutativity of Eq. (9), there are many different implementations of the same distribution of nodes. In particular, the prior in all nodes can be pooled separately for clarity, as was done for the full-field stimulus. Even if it is highly likely that statistical inferences are computed thanks to more complicated recurrent networks with local, lateral or feed-back connections, this model allows to understand a large class of non-linear systems as the simple interaction of linear local detectors and to test this simple hypothesis. We will now show that it applies to a large range of experiments involving spatio-temporal integration of motion information in primate OFR.

5. Spatial summation for OFR: effects of stimulus size

If we only consider a grating of speed $\vec{v}_g$ on a central disk aperture, it follows from Eq. (9) that the reference signal will be:

$$\vec{v}_m = \frac{\sum_{\mathbf{n} \in \mathscr{P}_c} \frac{1}{\sigma_x^2}}{\sum_{\mathbf{n} \in \mathscr{P}_c} \frac{1}{\sigma_x^2} + \sum_{\mathbf{n} \in \mathscr{P}} \frac{1}{\sigma_{p,x}^2}} \cdot \vec{v}_g \tag{11}$$

where $\mathscr{P}_c$ is the population of active central nodes (that is the intersection of $\mathscr{P}$ with the disk). It appears first that the prior may be simply pooled as $\sigma_p = (\sum_{\vec{x} \in \mathscr{P}} \frac{1}{\sigma_{p,\vec{x}}^2})^{-1/2}$.

Then, to model the integration over $\mathscr{P}_c$, we may consider that the density (or weight) of nodes pooling responses for the OFR is a centered Gaussian PDF of visual space with a width of ω. We may thus write that:

$$\sum_{\mathbf{n} \in \mathscr{P}_c} \frac{1}{\sigma_x^2} \propto C^2 \cdot \int_{0 \leqslant r \leqslant d} \exp\left(-\frac{r^2}{2 \cdot \omega^2}\right) \cdot 2 \cdot \pi \cdot r \cdot \mathrm{d}r$$

where d is the diameter and C the contrast. From

$$\int_{r=0}^{r=d} \exp\left(-\frac{r^2}{2 \cdot \omega^2}\right) \cdot 2 \cdot \pi \cdot r \cdot \mathrm{d}r$$
$$= 2 \cdot \pi \cdot \omega^2 \cdot \left(1 - \exp\left(-\frac{d^2}{2 \cdot \omega^2}\right)\right)$$

it follows that from Eq. (11),

$$\vec{v}_m(C, d) = \frac{1}{1 + \frac{C_e^2}{C^2 \cdot g^2(d)}} \cdot \vec{v}_g$$
$$\text{with } g^2(d) = 1 - \exp\left(-\frac{d^2}{2 \cdot \omega^2}\right) \tag{12}$$

where C_e is a constant corresponding to the half saturation contrast to the full-field grating. For a fixed diameter, the CRF will be again necessarily a Naka–Rushton curve (Naka and Rushton, 1966) of slope 2 with $C_{50} \propto g^{-1}(d) \cdot C_e$. This formula is to be compared with the
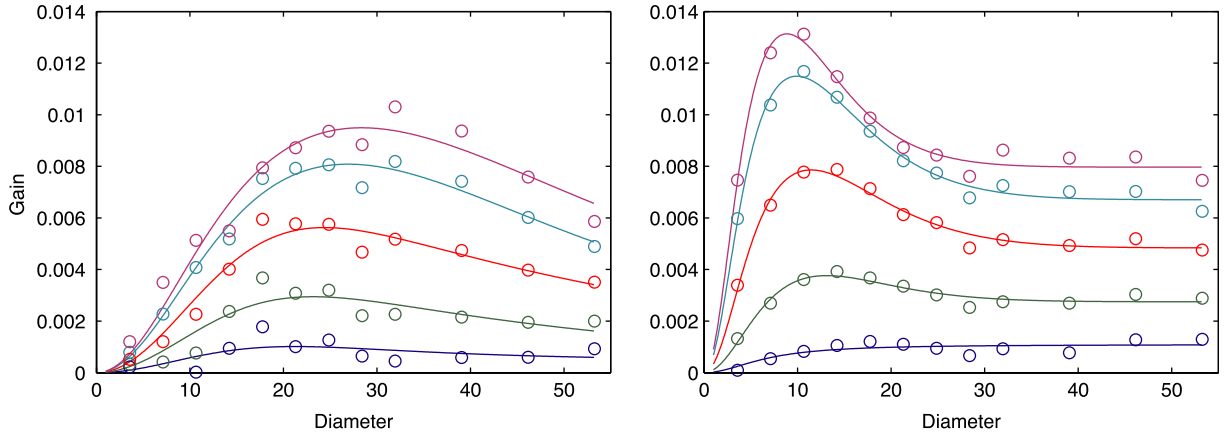
Fig. 5. Effects of size of a disk grating on the monkey's OFR response. We present here the gain of the oculo-motor response to a central grating (temporal frequency 10 Hz, see inset in Fig. 1a) as a function of its diameter for the macaque monkey (open circles) and the model (continuous line). Fits were performed as a function of the diameter d thanks to Eq. (13). (Left) At low frequencies (0.12 cpd) and contrasts, the gain increases monotonically with the diameter. The curves are well fitted by Eq. (12). (Right) However, in more general conditions (here 0.7 cpd), the initial gain decreases after a given diameter suggesting a suppressive effect. This corresponds to a surround inhibition which is well captured by Eq. (13), the inhibition being more pronounced when contrast is higher and contrary to intuition, the diameter ω extracted from the fits remains constant across curves.

Naka–Rushton gain, and shows that even if it similar qualitatively, the gain g^2 will give a better account of the integration on the disk.

Even if this family of curves shows a relatively good behavior for a large range of OFR responses (see low contrasts in Fig. 5, left) in particular for humans, they don't correspond to some observations, for instance for high grating frequencies (see Fig. 5, right). In fact, these responses show a suppression after a specific contrast (Sceniak et al., 1999), the so-called *super saturation*, which could not be accounted for this integration model. In fact, it appears clearly that the CRF will have the same variations as $g^2(d)$ and thus that it will be here necessarily monotonously increasing, i. e. the bigger the integration field's diameter, the higher the OFR gain.[4]

One solution is generally to add another integration term which accounts for this "surround suppression". As in the ratio of Gaussian (ROG) model of Cavanaugh et al. (2002), we may assume that the surround suppression is initiated by the pooling of information on $\mathscr{P}_c$ toward the null velocity[5] on a similar Gaussian distribution but with a larger size ω_i. It thus comes

$$\vec{v}_m = \frac{\vec{v}_g}{1 + \frac{\frac{C^2}{C_i^2} \cdot g_i^2(d)}{\frac{C^2}{C_e^2} \cdot g^2(d)}}$$

$$= \frac{\frac{C^2}{C_e^2} \cdot g^2(d) \cdot \vec{v}_g}{1 + \frac{C^2}{C_i^2} \cdot g_i^2(d) + \frac{C^2}{C_e^2} \cdot g^2(d)}$$

$$\text{with } g_i^2(d) = 1 - \exp\left(-\frac{d^2}{2 \cdot \omega_i^2}\right) \tag{13}$$

---

[4] This is congruous with the ideal observer model since for a given noise, more information should give a higher estimation of the translation speed.
[5] From Eq. (9), this formulation is quite general to explicit that there is a higher variance at higher eccentricities.

where C_i and ω_i are similar constants as in Eq. (12). This relation gives a parsimonious explanation of the ROG model by explicitly relating the division of the signal by an increase of pooled measurement variance. This equation provides a good fit of the behavioral data by tuning the parameters accordingly to a super-saturation (see Fig. 5, right). However, this model is still rather descriptive and further experiments have to be done to explicit the nature of this suppressive signal.

6. Center–surround integration in the bipartite stimulus

To challenge the possible origin of the surround suppression, we used the bipartite stimuli (see Fig. 6, top) for the OFR. It is constituted by a central grating as before but now surrounded by a perturbation of zero net velocity, the *flicker* stimulus. This stimulus is defined as the sum of a grating of speed $\vec{v}_f$ plus the same grating in the opposite direction:

$$I_{100}^f = \sin(2\pi f(x - \vec{v}_f \cdot t)) + \sin(2\pi f(x + \vec{v} \cdot t)) \tag{14}$$

$$= 2 \cdot \sin(2\pi fx) \cdot \sin(2\pi f \vec{v} \cdot t) \tag{15}$$

One can see that the corresponding likelihood is a gaussian centered on the null velocity stretched in the direction perpendicular to $\vec{v}_f$ (see Fig. 6, left). This signal with null velocity is similar to the one proposed in the previous section for the origin of the suppression. It thus comes that the integrative probability is a gaussian $\mathscr{N}(\vec{v}_m, \sigma_m)$ given by Eq. (9). It is possible to solve this equation and its analytical form takes a simpler form assuming that the variances perpendicular to the gratings are negligible. It comes for the vertical velocity:

$$\vec{v}_m(C) = \frac{1}{1 + \frac{C_g^2}{C^2}(\sigma_p^{-2} + \cos(\theta)\sigma_f^{-2})} \cdot \vec{v}_g \tag{16}$$
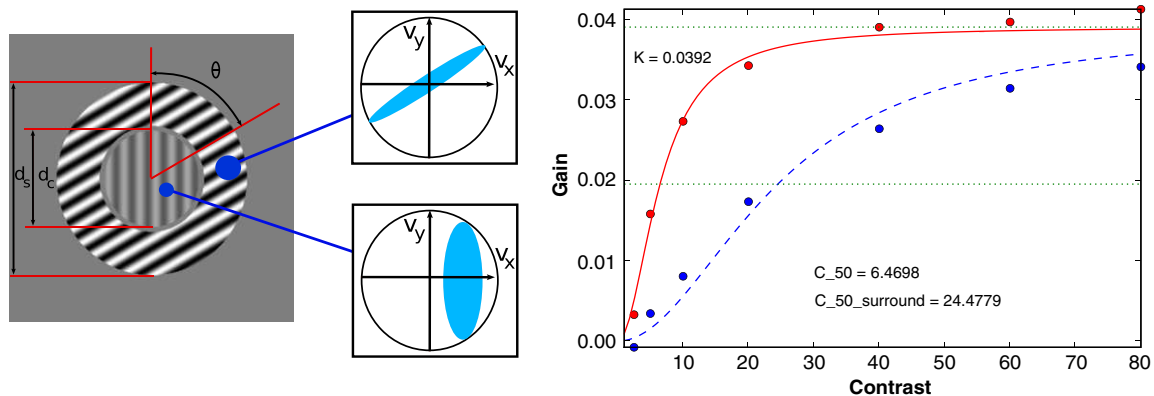
Fig. 6. Center–surround integration in the bipartite stimulus. (Left) To further evaluate the effect of the surround, the bipartite stimuli is defined by a central grating of diameter d_c, surrounded by an annulus of diameter d_s filled with a flicker stimulus of orientation θ. The flicker corresponds in a first approximation to a null velocity stimulus but from the aperture problem, it presents an elongated probability profile along this orientation. The grating varied in contrast, the probability getting broader as noise increased. (Right) Results show that suppression was dependent on the contrast of the grating and weak for a low flicker contrast. However, the suppression was stronger for a full contrast flicker and shows a dependance of the suppression as a function of the orientation of the surround as predicted by Eq. (16).

where θ is the angle of the flicker with respect to the central grating, σ_f the likelihood's standard deviation of the flicker, C the contrast of the central grating and σ_g its likelihood's standard deviation. One verifies that if the flicker is absent, $\sigma_f^{-2} = 0$ and the CRF is a standard slope-2 Naka–Rushton curve as in Eq. (7). A striking aspect of this relation is that if the contrast of the flicker increases (i.e. σ_f^{-1}), then the CRF will still be a slope-2 Naka–Rushton but that the only change is the contrast at half-saturation (as in Fig. 3, left): the curve will shift to a higher C_{50} value. This corresponds to the behavioral data (see the early and late responses in Fig. 1e) and was observed for various angles θ (see Fig. 6, right).

7. Discussion: beyond the behavioral receptive field

We proposed here a simple formulation for the integration of elementary information (see Eq. (9)) that we applied successfully to oculo-motor data for OFR in the primate (see Figs. 5 and 6). It is based on the normality of the distribution at the different nodes but more importantly on the hypothesis of independence for the measurement noise of the different nodes. This model thus gives a simple model for the function of spatio-temporal integration in the visual system as an ideal observer and gives in particular an account for both subtractive or divisive mechanisms by relating them respectively to a change across nodes in $\vec{v}_n$ or in variance σ_n^2. This model is similar to different accounts on the contrast gain control that appears to be a main feature of low-level visual areas. First, it is similar to divisive normalization (Schwartz and Simoncelli, 2001), since it uses a similar method for computing probabilities, except that our formulation is more general in terms of the qualitative interpretation of the spatio-temporal integration. Also, this model could be implemented in a simple manner using a spiking neural network (Perrinet, 2004; Perrinet

et al., 2005) similarly to the feed-forward linear/non-linear model (Carandini et al., 2005). However, in our case the non-linearity will be explicitly defined by the nodes' probabilities and not be set arbitrarily. Finally, this model gives an account for the results of Cavanaugh et al. (2002) and are compatible with the ratio of Gaussian (ROG) model by stating that the "read-out" from the OFR could be seen as the interaction of two integration fields with different sizes and polarities. In particular, the introduction of the surround field is more a descriptive account of the data and its characteristics may change with the stimulus' properties. However, we gave a possible explanation in probabilistic terms which may be induced by an increase of noise with eccentricity.

Nonetheless, the results in the first experiment showing a super-saturation behavior (e.g. at high frequencies) are intriguing (see Fig. 5): for a given noise level, after a certain diameter, integrating on a larger field leads to a decrease of the response gain. This is in contradiction with the ideal observer under a general hypothesis, which should always give a higher gain when more information is present in the image. To overcome the hypothesis of the existence of a "suppressive field" as above, which would be more descriptive than explanatory, we propose that the main hypothesis of the ideal observer still holds, but that for these particular stimuli, the information is coded in a different way. First, we will explore in future work the influence of line endings which may have a significant relative importance, but also the possibility that nodes may propagate information laterally. As a matter of fact, most natural motions are rigid and knowing the speed with accuracy at a certain time, one could predict the accuracy at a latter time at the predicted change of location. This would break down the hypothesis of independence that led to Eq. (9) – where all nodes may be spatially permuted with a random shuffle – and could account for the fact that for gratings

with big diameters the mistuning of responses in space, due to the different selectivity of neurons as a function of eccentricity, would lead to a decrease of total gain.

Reproducible research

Scripts reproducing all figures may be obtained from the author upon request and on the author's web-site at http://incm.cnrs-mrs.fr/LaurentPerrinet.

References

Albrecht, D., Geisler, W., Frazor, R., Crane, A., 2002. Visual cortex neurons of monkeys and cats: temporal dynamics of the contrast response function. Journal of Neurophysiology 88, 888–913.

Barthélemy, F., Masson, G.S., 2006. Spatial integration of motion for human and monkey ocular following: effects of spatial frequency and eccentricity. In: Society for Neurosciences Abstracts, number 735.5/L3.

Barthélemy, F.V., Vanzetta, I., Masson, G.S., 2006. Behavioral receptive field for ocular following in humans: dynamics of spatial summation and center–surround interactions. Journal of Neurophysiology (95), 3712–3726. doi:10.1152/jn.00112.2006.

Barthélemy, F., Perrinet, L.U., Castet, E., Masson, G.S., 2007. Dynamics of distributed 1D and 2D motion representations for short-latency ocular following. Vision Research, in press, doi:10.1016/j.visres.2007.10.020.

Born, R., Tootell, R., 1991. Single-unit and 2-deoxyglucose studies of side inhibition in macaque striate cortex. Proceedings of the National Academy of Sciences USA 88, 7071–7075.

Carandini, M., Demb, J.B., Mante, V., Tolhurst, D.J., Dan, Y., Olshausen, B.A., Gallant, J.L., Rust, N.C., 2005. Do we know what the early visual system does? Journal of Neuroscience 25 (46), 10577–10597. doi:10.1523/JNEUROSCI.3726-05.2005.

Cavanaugh, J.R., Bair, W., Movshon, A., 2002. Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. Journal of Neurophysiology 88 (5), 2530–2546. doi:10.1152/jn.00692.2001.

Duffy, C., Wurtz, R., 1991. Sensitivity of MST neurons to optic flow stimuli. i. A continuum of response selectivity. Journal of Neurophysiology 65, 1346–1359.

Eifuku, S., Wurtz, R., 1999. Response to motion in extrastriate area MST: disparity sensitivity. Journal of Neurophysiology 82, 2462–2475.

Gellman, R., Carl, J., Miles, F., 1990. Short-latency ocular-following responses in man. Visual Neuroscience 5, 107–122.

Heuer, H., Britten, K., 2002. Contrast dependence of response normalization in area MT of the rhesus macaque. Journal of Neurophysiology 88, 3398–3408.

Hurlimann, F., Kiper, D.C., Carandini, M., 2002. Testing the bayesian model of perceived speed. Vision Research 42 (19), 2253–2257.

Kawano, K., 1999. Ocular tracking: behavior and neurophysiology. Current Opinion in Neurobiology 9, 467–473.

Kawano, K., Shidara, M., Watanabe, Y., Yamane, S., 1994. Neural activity in cortical area MST of alert monkey during ocular following responses. Journal of Neurophysiology 71, 2305–2324.

Lisberger, S.G., Morris, E.J., Tychsen, L., 1987. Visual motion processing and sensory-motor integration for smooth pursuit eye movements.

Annual Review of Neuroscience 10, 97–129. doi:10.1146/annurev.ne.10.030187.000525.

Masson, G.S., 2004. From 1D to 2D via 3D: surface motion integration for gaze stabilization in primates. Journal of Physiology (Paris) 98, 35–52.

Masson, G.S., Castet, E., 2002. Parallel motion processing for the initiation of short-latency ocular following in humans. Journal of Neuroscience 22 (12), 5149–5163.

Masson, G.S., Rybarczyk, Y., Castet, E., Mestre, D.R., 2000. Temporal dynamics of motion integration for the initiation of tracking eye movements at ultra-short latencies. Visual Neuroscience 17 (5), 753–767.

Masson, G.S., Yang, D., Miles, F., 2002. Reversed short-latency ocular following. Vision Research 42 (17), 2081–2087.

Miles, F.A., Kawano, K., Optican, L.M., 1986. Short-latency ocular following responses of monkey. i. Dependence on temporospatial properties of visual input. Journal of Neurophysiology 56 (5), 1321–1354.

Naka, K.I., Rushton, W.A., 1966. S-potentials from luminosity units in the retina of fish (Cyprinidae). Journal of Physiology 185 (3), 587–599.

Nowlan, S.J., Sejnowski, T.J., 1995. A selection model for motion processing in area MT of primates. Journal of Neuroscience 15, 1195–1214.

Osborne, L., Bialek, W., Lisberger, S., 2004. Time course of information about motion direction in visual area MT of macaque monkeys. Journal of Neuroscience 24, 3210–3222.

Pack, C., Born, R., 2001. Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. Nature 409, 1040–1042.

Perrinet, L.U., 2004. Feature detection using spikes: the greedy approach. Journal of Physiology (Paris) 98 (4-6), 530–539. doi:10.1016/j.jphysparis.2005.09.012.

Perrinet, L.U., Barthélemy, F., Castet, E., Masson, G.S., 2005. Dynamics of motion representation in short-latency ocular following: a two-pathways bayesian model. In: Carmona, R.A., Linan-Cembrano, G., (Eds.), Perception, ECVP, vol. 34. p. 38.

Priebe, N., Lisberger, S., 2004. Estimating target speed from the population response in visual area MT. Journal of Neuroscience 24 (8), 1907–1916.

Reynaud, A., Barthélemy, F., Vanzetta, I., Masson, G.S., Chavane, F., 2007. Input–output transformation in the visuo-oculomotor loop: comparison of real-time optical imaging recording in V1 to the ocular following response to center–surround stimulation. Italian Archives of Biology 145, 251–262.

Sceniak, M., Ringach, D., Hawken, M., Shapley, S., 1999. Contrast's effect on spatial summation by macaque V1 neurons. Nature Neuroscience 2 (August), 753–759.

Schwartz, O., Simoncelli, E., 2001. Natural signal statistics and sensory gain control. Nature Neuroscience 4 (8), 819–825.

Sclar, G., Maunsell, J., Lennie, P., 1990. Coding of image contrast in central visual pathways of the macaque monkey. Vision Research 30, 1–10.

Sheliga, B., Chen, K., EJ, F., Miles, F., 2005. Ocular following in humans: a response to first-order motion energy. Vision Research 45, 3307–3321.

Simoncelli, E.P., Heeger, D.J., 2001. Representing retinal image speed in visual cortex. Nature Neuroscience 4, 461–462. doi:10.1038/87408.

Smith, M., Majaj, N., Movshon, J., 2005. Dynamics of motion signaling by neurons in macaque area MT. Nature Neuroscience 8, 220–228.

Takemura, A., Kawano, K., 2002. Sensori-to-motor processing of the ocular following response. Neuroscience Research 43, 201–206.

Takemura, A., Murata, Y., Kawano, K., Miles, F.A., 2007. Deficits in short-latency tracking eye movements after chemical lesions in monkey cortical areas MT and MST. Journal of Neuroscience 27 (3), 529–541. doi:10.1523/JNEUROSCI.3455-06.2007.

Weiss, Y., Simoncelli, E.P., Adelson, E.H., 2002. Motion illusions as optimal percepts. Nature Neuroscience 5 (6), 598–604. doi:10.1038/nn858.