Network: Computation in Neural Systems
September 2008; 19(3): 183–212

informa
healthcare

Efficient encoding of natural optic flow

DIRK CALOW & MARKUS LAPPE

Department of Psychology, Westfälische Wilhelms University, Münster, Germany

(Received 17 March 2008; revised 20 May 2008; accepted 25 July 2008)

Abstract
Statistically efficient processing schemes focus the resources of a signal processing system on the range of statistically probable signals. Relying on the statistical properties of retinal motion signals during ego-motion we propose a nonlinear processing scheme for retinal flow. It maximizes the mutual information between the visual input and its neural representation, and distributes the processing load uniformly over the neural resources. We derive predictions for the receptive fields of motion sensitive neurons in the velocity space. The properties of the receptive fields are tightly connected to their position in the visual field, and to their preferred retinal velocity. The velocity tuning properties show characteristics of properties of neurons in the motion processing pathway of the primate brain.

Keywords: *Optic flow, natural ego-motion, statistics, mutual information maximization, encoding*

Introduction

Although the processing power of the brain is huge compared with contemporary artificial signal processing systems the range of signals the brain can process is limited. The visual pathways of the brain show adaptations to the statistics of the natural environment for an efficient processing of the set of signals that the environment provides. Such adaptations are seen in gestalt laws (Elder and Goldberg 2002; Krüger and Wörgötter 2002) and in efficient encoding schemes (Barlow 1961; Laughlin 1981; Linsker 1988; Atick 1992) in which the processing pathway is more sensitive for signals that occur very frequently than to signals that are very unlikely to occur. However, in many natural situations the visual input is dynamic because animals move. We aim here to apply the concept of efficient

Correspondence: D. Calow, Department of Psychology, Westfälische Wilhelms University, Fliednerstrasse 21, Münster 48149, Germany. Tel: 00492518334171. Fax: 004992518334173. E-mail: calow@uni-muenster.de

ISSN 0954-898X print/ISSN 1361-6536 online/08/030183–212 © 2008 Informa UK Ltd.
DOI: 10.1080/09548980802368764

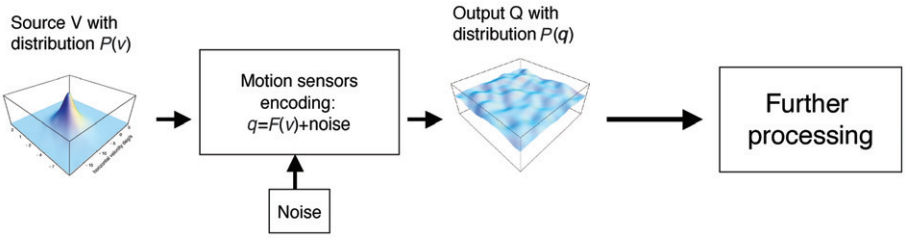

Figure 1. Scheme of a channel.

encoding to the realm of motion processing to find properties of motion sensitive neurons, which efficiently encode the set of motion signals generated on the retina by natural behavior in natural environments. Most of the retinal motion in natural situations is generated by ego-motion of the observer. Therefore, we concentrate our investigation on retinal motion signals during ego-motion.

The concept of efficient encoding is to maximize the mutual information between input and output. The mutual information describes how tightly an output is statistically linked to an input. Input and output are related by an encoding scheme or channel, which is composed of a set of encoding units, or neurons (Figure 1). The input signal, or source, activates the sensors and elicits a set of signals that the channel can convey. In the case of visual motion the input signal v is the spatial-temporal luminance change of the retinal image, and the sensors are neurons in the motion processing pathway of the brain. The encoding scheme imposes a transformation on the input signal, which is expressed by a transformation function F, such that the encoded signal q equals $F(v)$ plus a noise term, which is assumed to reflect noise in the physical and chemical processes in the brain. Eventually, the outcome of the encoding is an output signal which is linked to the input signal and which is the basis for further processing and behavior. The properties of the relationship between input and output are governed by the transformation function F and the properties and strength of the noise.

In this study, we search for transformation functions F that are adapted to the local statistics of retinal motion signals such that mutual information between retinal motion and the encoded signals is maximized for each position in the visual field. On the basis of F we construct for each position in the visual field a set of efficient retinal motion encoding units, and analyze their direction and speed tuning properties. These tuning properties depend on the position in the visual field and can serve as predictions for the retinal motion sensitivities and tuning properties along the motion processing pathway of the brain. At the outset, this view on efficient encoding relies exclusively on the statistics of motion signals and considers all retinal positions, motion direction, and speeds of equal importance. One can imagine that some motion signals are more important for the behavior of the system than others and that errors in detecting such motion signals would have harsher consequences. Earlier work (Calow and Lappe 2007) has indeed shown that some parts of the visual field carry more information about heading whereas others carry more information about environment layout. However, including these aspects is beyond the scope of the present study since it would require different encodings for different task objectives. Our approach assumes a single representation that attempts

to most faithfully and most efficiently encode the motion signals for any later task-specific processing.

The analysis of the present examination is relying on the statistics of true retinal motion calculated from geometrical optics rather than motion estimated by early motion estimators from spatial-temporal image luminance variations (Calow et al. 2004; Calow and Lappe 2007; Roth and Black 2007). We generate true retinal motion fields based on the Brown range image database (Huang 2000) in combination with biologically plausible ego-motion. In natural locomotion, eye movement reflexes stabilize gaze on objects of interest in the scene (Solomon and Cohen 1992; Lappe et al. 1997; Niemann et al. 1999) such that natural ego-motion is always a combination of body movement and eye movement. The combination of body movement, eye movement, and depth structure of the visual environment determines the structure of the optic flow on the retina (Lappe et al. 1999). Without denying the importance of results relying on real camera motion (Betsch et al. 2004; Zanker and Zeil 2005) we see several advantages concerning our approach. First, our procedure makes the results independent of the properties of particular motion detectors and their notorious performance problems (Barron et al. 1994). Second, our method provides a huge number of statistically-independent retinal motion fields for a variety of human imitating ego-motion situations and natural scenes that would be difficult to obtain from camera motion which imitates human like ego-motion including gaze stabilization. Third, our approach provides the ground truth on which the statistics rely and has therefore the potential to investigate higher function of motion processing like heading detection, motion in depth and figure ground segregation from motion. However, while we believe this approach to be valuable we also acknowledge the necessity to investigate the statistics of retinal motion as generated by spatial-temporal luminance changes and evaluated by early motion sensors. Yet, in this study we will focus on the investigation of true motion fields.

In the motion processing pathway of the primate brain, visual motion selectivity is first established in area V1 and then routed via the middle temporal (MT) area to areas MST and VIP, which analyze optic flow for the estimation of self-motion (Lappe 2000). In this pathway, area MT provides a representation of the optic flow field that is well suited as input for optic flow analysis. We suggest that area MT is a likely candidate for an efficient encoding scheme for optic flow for several reasons. First, area MT is the first stage along the motion processing pathway that extract retinal velocity rather than merely spatial-temporal frequency (Perrone and Thiele 2001). Therefore, the responses of neurons in area MT can be regarded as indicators for true motion sensitivity. Second, the receptive field sizes of MT neurons are well adapted to construct a robust, noise-reduced representation of the optic flow from local motion inputs (Calow et al. 2005) which is close to the true motion. Third, higher brain areas like MST and VIP, which receive their input from area MT, already process more complex information from optic flow, such as the pattern of flow (Saito et al. 1986; Duffy and Wurtz 1991; Orban et al. 1992) or the direction of heading (Lappe et al. 1996). Neurons in these areas must have a complex substructure of the receptive field in order to perform calculation on flow patterns (Beintema et al. 2004). However, since the present study is relying on the local statistics of retinal motion (Calow and Lappe 2007), the resulting efficient encoding units are simple motion sensors. They have no spatial extension and

therefore no spatial receptive field structure. They respond simply to speed and direction and have no special response behavior for motion patterns. Investigations of pattern selectivity requires the use of second-order statistical properties of retinal motion which is a question for future research.

Even though the efficient detectors that we construct here have no extended receptive fields, we argue that the proposed scheme leads to the construction of motion sensitive neurons which reflects some of the aspects of the motion representation in area MT, and, moreover, leads to predictions about further properties of neuron populations concerning their tuning properties and their characteristics. Similarities between neuron properties found in area MT and properties of motion sensitive units constructed in this study suggest in turn that the motion processing pathway of the brain might have evolved an efficient encoding scheme.

The present study continues our previous investigation of the local statistics of optic flow for self motion through natural scenes (Calow and Lappe 2007). That previous investigation established methods to analyze the statistical properties of retinal motion. The present study relies on these methods but needs some modifications in measuring the statistics of optic flow. Thus, before explaining how the maximizing of mutual information between input and output is achieved, and how the transformation function F is constructed, we will briefly explain how to construct the prior distributions of retinal motion signals which form the database for the construction of efficient retinal motion encoding units. Thereafter we construct the efficient retinal motion encoding units for each position in the visual field. Finally, we evaluate the properties of these units and interpret and discuss them.

Methods

Construction of retinal flow fields

To construct the database of retinal flow fields we largely follow the method described in (Calow and Lappe 2007). Briefly, the calculation of retinal optic flow fields relies on the knowledge of the depth map of a variety of natural scenes, which comprises urban and forest scenes. We use the Brown Range Image Database, a database of 197 range images collected by Ann Lee, Jinggang Huang and David Mumford at Brown University (Huang 2000). The range images were recorded with a laser range-finder with high spatial resolution (Figure 2). The knowledge of the 3D coordinates of each image point allows the calculation of the true motion of that point for any given combination of translation and rotation of the projection surface (i.e., the retina of the observer). As we are interested in the statistics of retinal projections, we consider the retina as a section of a spherical projection surface with unit radius. The next step then involves finding a set of ego-motion parameters to construct flow fields from the range data.

Ego-motion parameters. **Heading, eye rotation and the distribution of walking speed** To calculate the flow field from the scene structure we need the motion parameters of the projection surface. The ego-motion of the surface is fully

Figure 2. Panoramic projection of 3D data of a range-image, The grey values encode the intensity of the reflected laser beam.

described by the translational velocity vector of the surface $T = (T_x, T_y, T_z)$ and the vector of rotation $\Omega = (\Omega_x, \Omega_y, \Omega_z)$ in the Euclidian coordinate system attached to the projection surface. The translation T of the surface can be further split into the speed $\|T\|$ and heading (H_ϕ, H_θ), which are azimuth and elevation denoting the direction of the translational velocity vector of the surface:

$$T = (T_x, T_y, T_z) = \|T\|(\cos(H_\phi)\sin(H_\theta),\ \cos(H_\theta),\ \sin(H_\phi)\sin(H_\theta)).$$

Natural ego-motion within the scenes involves eye movements which stabilize the gaze on environmental objects (Lappe et al. 1998; Lappe and Hoffmann 2000). Gaze stabilization keeps the point of interest or the gaze attracting object in the center of the visual field and causes the motion in the center of view to be zero. The associated rotation depends on the translation by $\Omega = (1/Z_f)(Ty, -T_x, 0)$, where Z_f denotes the depth of the point at which gaze is directed. But gaze stabilization is not perfect. The ratio between stabilizing eye rotation and retinal speed at the fovea is found to be broadly distributed around a peak of 0.5 not effected by observer speed (Lappe et al. 1998). Accordingly, we introduce a gaussian distributed stabilizing factor S, such that the stabilizing eye rotations are $\Omega = (S/Z_f)(Ty, -T_x, 0)$. The mean of the distribution of S is 0.5 and the standard deviation is 0.5.

We also have to take into account that walking speed can range from strolling very slowly to fast walking and running. To incorporate this wide range of possible walking speeds in our considerations, we introduce the following distribution governing the walking speed:

$$P(\|T\|) = \frac{1}{\sigma\|T\|\sqrt{2\pi}}\exp(-(\log(\|T\|) - \mu)^2/2\sigma^2),$$

where $\sigma = 0.6$ and $\mu = \log(1.4) + \sigma^2$. This logarithmic gaussian distribution has the peak at $1.4\,\mathrm{m\,s^{-1}}$, which is considered as the usual walking speed. The parameter $\sigma = 0.6$ is designed such that a sprint with a speed higher than $8\,\mathrm{m\,s^{-1}}$ is very unlikely ($\int_{8\,\mathrm{m\,s^{-1}}}^{\infty} P(x)\mathrm{d}x = 0.01$). To consider all aspects of human walking, we also take bouncing and swaying of the head during walking, into account (Imai et al. 2001; Calow and Lappe 2007).

To determine possible walking directions within a range image we search for areas which are free from obstacles in a depth of at least 3 m and a width of 0.7 m. This criterion gives us a set of walking directions for each scene, which are considered to be equally likely.

In natural locomotion, the actual combination of walking direction, walking speed, gaze direction and head movement is linked to the environment and to the task. Moreover, the simulated components of ego-motion might not be in general independent. Therefore, our assumed ego-motion must be regarded as an approximation to actual ego-motion and might not match actual ego-motion in all details. However, our simulation matches the main components of human ego-motion and allows us to combine naturalistic ego-motion parameters with the true depth information data provided by the range image database. Alternatives for generating input datasets, such as moving a camera through natural environments (Betsch et al. 2004; Zanker and Zeil 2005) would not allow to capture the scene depth structure and calculate true retinal motion fields.

Gaze directions The retinal flow depends on the heading with respect to the direction of gaze. To obtain a distribution of gaze directions we measured eye and head positions of observers who walked straight ahead through forest and urban environments similar to the scenes recorded by the laser range finder. Simultaneously measuring eye and head positions allows to calculate the gaze directions in the scene which depends on the position on the head and the position of the eye within the head fixed coordinate system. The positions of the eyes were measured with a portable eye-tracking system (View Point eye-tracker from ArringtonResearch). The system is equipped with infrared cameras for tracking the position of the pupils, infrared light emitting diodes for illuminating the eyes and a camera on the forehead which records the scene. The cameras and LEDs are mounted on a frame which can be worn like glasses (Figure 3a). Head positions were measured by a 3° of freedom orientation tracking system (inertia cube 2 from Intersense, Figure 3b). This device was fixed on a light helmet worn by the subject. Both, the eye tracker and the inertia cube must be calibrated and mutually aligned. The eye tracker was calibrated in the coordinate system of the scene camera. During the calibration, the subject has to gaze toward nine subsequent view directions arranged as a grid in the camera image. After the calibration the view directions of the subject in relation to the head camera were known. To align the head camera with the inertia cube, the subject was asked to fixate a reference point which was presented in the direction

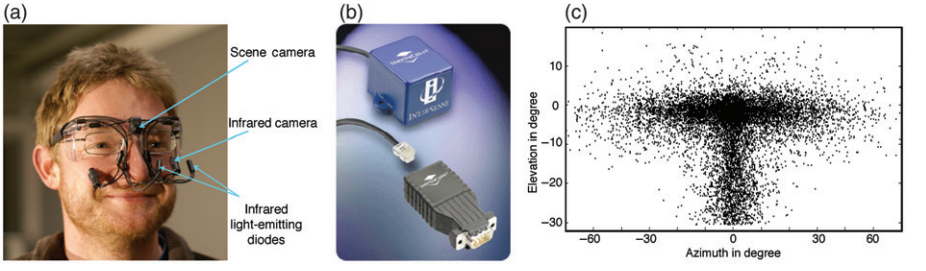

Figure 3. (a) Portable eye-tracker for measuring view directions during walking. The position of the eye, which is illuminated by an infrared light-emitting diode, is detected by an infrared camera. The scene is simultaneously recorded by a scene camera. (b) The orientation tracking system inertia cube 2 (Intersense) is used to measure head orientation. (c) Measured distribution of gaze directions while walking in relationship to the walking direction.

of walking at eye level. This allowed for aligning the coordinate system of the inertia cube with the scene camera. The combination of head orientation and view direction in relation to the camera gives the absolute gaze direction in the world during walking. From the entire set of absolute gaze directions all phases of pursuit and fixation were extracted. Figure 3(b) shows the set of gaze directions used for the further analysis. The distribution of gaze directions shows two clusters:, a vertically aligned cluster related to the observation of the path in front of the subject, and a horizontally aligned cluster related to scanning the scene and gazing toward objects.

A detailed analysis of view directions while walking including the analysis of saccades, velocity of smooth pursuit and fixations will be presented elsewhere. For the present study, we combine each walking direction within the natural scenes provided by the laser range finder with a set of 30 randomly selected gaze directions calculated from the measured distribution in terms of a uniforming procedure (Appendix, Section 'Generating a new random sample from the transformation function F_Ψ').

The retina. The field of view of the circular retina is set to 90° horizontally and 20° vertically. This field of view is subdivided in pixels, with a resolution of $0.36^\circ \times 0.36^\circ$ yielding a grid of 250×160 pixels. As the angular separation of the range images is 0.18°, one pixel covers up to four data points. The depth values $Z = R\sin(\phi)\sin(\theta)$ from these data points are averaged. The mean depth value is assigned to the pixel in question. Thus, we reduce the original resolution provided by the laser range images. This procedure is necessary, because the pixel grid of the retina is sliding over the pixel grid of the laser range image and mostly does not match the original pixels. Therefore, reducing the resolution ensures that all pixels of the retina receive appropriate motions signals.

We constructed around 21000 different flow fields. Each flow field represents a snap-shot of motion signals for a certain ego-motion situation. This gives a distribution of around 21000 sample motion signals for each position of the considered visual field. Except for regions that present sky in the image, each considered position in the visual field provide a flow vector. Figure 4 shows some distributions of motion signals for different positions in the visual field.

Maximizing mutual information

In this section we generally explain, how the maximizing of mutual information between input and output is achieved, and how the transformation function F is constructed. Let us assume that the statistics of the input signals $s \in \mathbb{R}^2$ follow a certain probability density function (PDF) $P(s)$. Furthermore, suppose that the range of output signals that the channel can provide is limited. If the range of input signals is unlimited the transformation function must bend the input signals at least at the tails of the PDF, such that the complete dispersion of the input can be confined to the range of the output signals. Considering the entire set of possible input signals the mutual information between the distributions of output and input signals is maximal when knowing the output provides the maximal information about the input. In the following we will say that a distribution of signals is efficiently

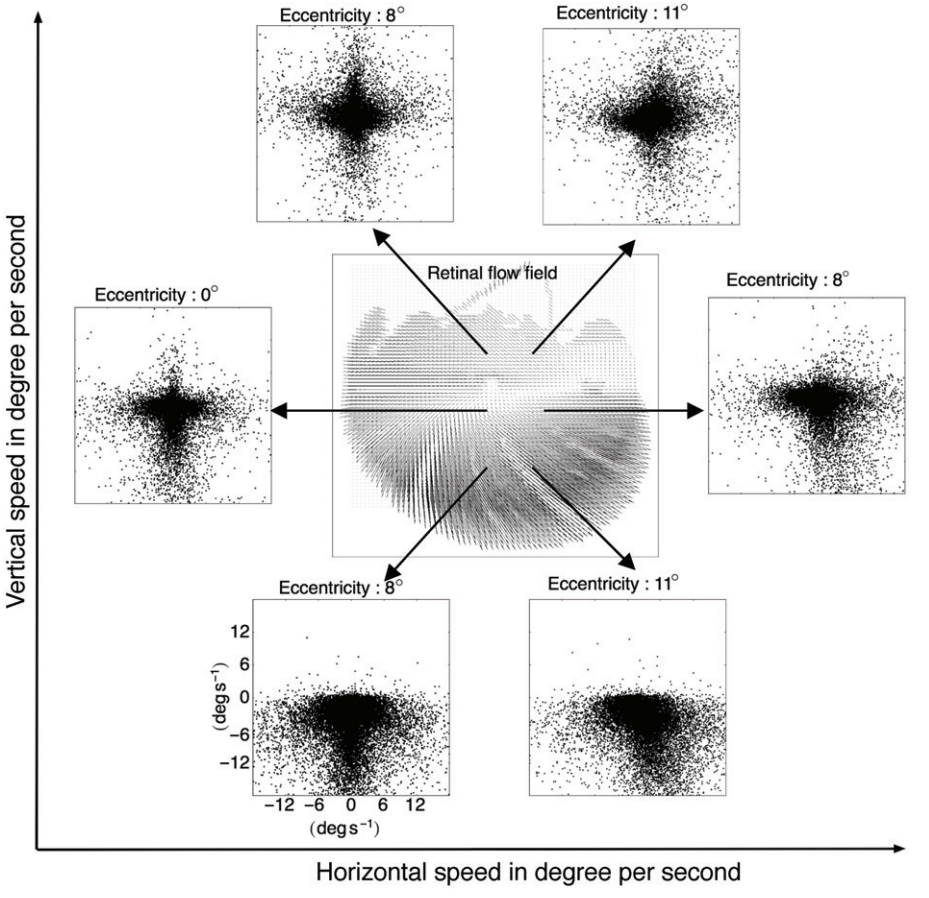

Figure 4. Scatterplots of distributions of actual retinal motion signals for some example positions in the field of view, depicted in cartesian coordinates. Each dot represents a motion signal at the respective view field position for a certain ego-motion situation. Such datasets are acquired for each considered position in the visual field and form the basis for the statistical analysis.

encoded if the mutual information between the output distribution and the input is maximal. This assumption requires a sophisticated transformation function, the properties of which depend on the shape of the input PDF. Following the derivation in the Appendix, Section 'Maximizing mutual information: Derivation of the transformation function', we use the following transformation function $F: \mathbb{R}^2 \rightarrow [0, 1] \times [0, 1]$; $F(s) = (f^1(s^1, s^2), f^2(s^1, s^2))$ for maximizing the mutual information between input and output:

$$f^1(s^1, s^2) = f^1(s^1) = \int_{-\infty}^{\infty} H_s(s^1 - t^1) P(t^1) \mathrm{d}t^1 \tag{1}$$

$$f^2(s^1, s^2) = \int_{-\infty}^{\infty} H_s(s^2 - t^2) P(t^2|s^1) \mathrm{d}t^2, \tag{2}$$

where H_s is the Heaviside step function. In the following, we will refer to this transformation and to its approximated application to datasets (Appendix, Section 'Uniforming of datasets and the approximation of the transformation function') as uniforming.

The transformation functions described by Equations 1 and 2 are in fact the cumulative functions for the PDF $P(s)$. The transformation function belongs to a class of bounded transformation functions that maximize information transfer following the infomax principle of Linsker (1988) see also (Laughlin 1981; Atick 1992) for two-dimensional input. While Section 'Maximizing mutual information: Derivation of the transformation function' of the Appendix provides a detailed derivation, it is intuitively clear that the transformation function maximizes the entropy of the transformed output signal and generates a factor code. The resulting random variables are statistically independent, i.e., $P(f^1, f^2) = P(f^1)P(f^2)$.

Setting the image of the transformation function F to $[0, 1] \times [0, 1]$ does not imply a loss of generality but accounts for the fact that the processing capacity and the output range of the channel are bounded. In the brain, the output range of early signal processing is bounded simply because a finite number of neurons are involved. Topologically, each confined two-dimensional output space can be one to one mapped to $[0, 1] \times [0, 1]$.

Polar representation of retinal flow vectors. The problem of efficient encoding by maximizing the mutual information between input and output can be solved by different encoding schemes and different transformation functions, which are mathematically identical but may lead to different representations. One has to choose a particular representation, however, to perform the analysis. The random variables we choose for the further analysis of the processing of retinal flow are retinal speed V and retinal direction Φ. Retinal direction is calculated as the deviation from the radial direction. The choice of polar coordinates is justified for the following reasons. Direction and speed are already largely independent for natural retinal flow (Calow and Lappe 2007) and carry different information about the scene and ego-motion. Because this is a property of the flow statistics, any efficient encoding scheme will inherit this property, which is then reflected in the properties of the encoding units or receptive fields. Thus, the MT-like independence of direction and speed (Rodman and Albright 1987) is indeed a prediction from efficient encoding given by the statistics of the optic flow and is not related to our choice of a coordinate system. However, choosing this coordinate system of the random variables eases the calculations, saves de-correlation effort, and gives direct access to the information in optic flow. Thus, polar coordinates rather than, for example, cartesian coordinates are the appropriate choice to investigate their properties.

The construction of efficiently encoding neuron populations

Figure 5(a), shows the distribution of motion signals for a position at 8° eccentricity (depicted in Figure 4, right panel) in polar coordinates. The horizontal axis denotes the retinal direction Φ in degrees. The vertical axis denotes retinal speed V plotted on a logarithmic scale. The first step in the construction of an efficient

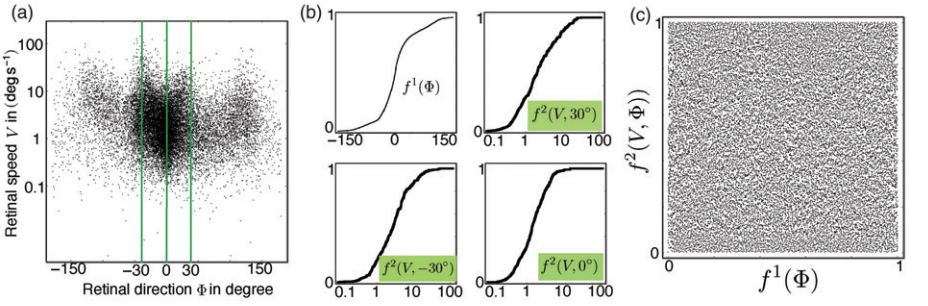

Figure 5. Uniforming procedure for a retinal motion distribution on the basis of an example dataset acquired at a visual field position at 8° (Figure 4 right panel), (a) The dataset pictured in Figure 4 right panel transformed in polar coordinates direction Φ and speed V, where the latter is plotted on a logarithmic scale. $\Phi = 0$ corresponds to radial direction. (b) The approximated transformation function $F: \mathbb{R} \to [0,1] \times [0,1]$, $f^1(\Phi)$ transforms direction, three examples of $f^1(V, \Phi)$ for $-30°$, 0 and $30°$ are shown. The green lines in (a) mark the angles where the example functions act. (c) The resulting uniform distribution $(f^1(\Phi_i), f^2(V_i, \Phi_i))$.

encoding of this motion dataset is to apply Equations 15 and 16 (Appendix) to the dataset $\{\Phi_i, V_i\}_{i=1,\ldots,N}$. This gives the approximation of the transformation function F (Equations 1 and 2). Figure 5(b) shows the first component $f^1(\Phi)$ and three sections $f^2(\Phi, V)$ at -30, 0 and $30°$. The result of the mapping of the dataset by F can be seen in of Figure 5(c). The resulting distribution is a uniform distribution in the space $[0,1] \times [0,1]$ (χ^2-test, $\chi^2 < 3.0$ for $24°$ of freedom (5×5 bins), $p > 0.99$ for all transformed distributions).

The above procedure constructs the transformation function F. To construct the units which efficiently encode the input signal and which perform the transformation we cover the resulting uniform distribution with circular gaussian receptive fields (Figure 6a). In the following we will refer to these units as neurons. The mean response r_m of such a neuron elicited by an input signal (Φ, V) is then described by

$$r_m(\Phi, V) = r_0 \exp\left(-\frac{|2\pi(c_1 - f^1(\Phi))|^2_{\text{circ}}}{2(2\pi w)^2}\right) \exp\left(-\frac{(c_2 - f^2(\Phi, V))^2}{2w^2}\right). \qquad (3)$$

In Equation 3, $(c_1, c_2) \in [0,1] \times [0,1]$ marks the center of the receptive field, w steers the width, and the notation $|\cdots|_{\text{circ}}$ means

$$|x|_{\text{circ}} = \begin{cases} |x|; & |x| < \pi \\ 2\pi - |x|; & |x| \geq \pi \end{cases}; \quad x \in [0, 2\pi].$$

The latter is the metric on a circle with radius 1 and accounts for the fact that the space of directions Φ is compact and that the points $-\pi$ and π are identical. In the following we fix the width at $w = 0.1$ by an arbitrary choice which produces reasonable results.

Figure 6(b) shows the mean response behavior of four example neurons in the original velocity space. The circles denote the values for retinal speed in a logarithmic scale and the lines denote the direction. Thus, by uniformly covering

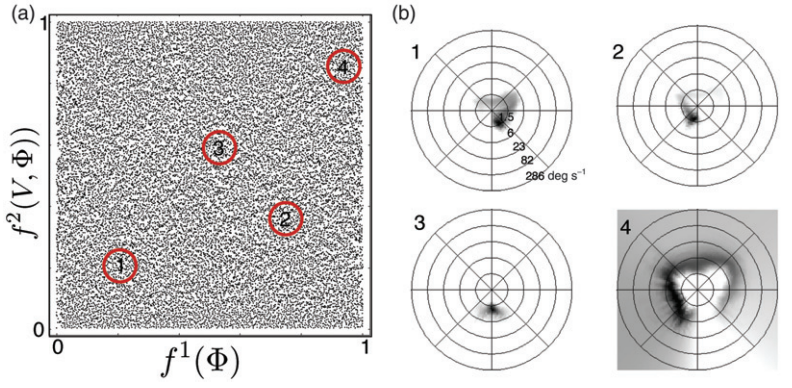

Figure 6. (a) The distribution $(f^1(\Phi_i), f^2(V_i, \Phi_i))$ resulting from uniforming (Figure 5). Circles mark the positions and receptive fields of four example neurons drawn from the neuron population, the receptive fields of which uniformly cover the distribution $(f^1(\Phi_i), f^2(V_i, \Phi_i))$. (b) Plots of velocity sensitivity of the four example neurons. Black means high response, white marks vanishing response.

the space $[0, 1] \times [0, 1]$ by a grid of such neurons we obtain a neuron population which turns the original signal distribution into a uniform distribution. How many neurons are necessary to effectively encode the original distribution? This question cannot be answered quantitatively but it appears evident that a distribution with high variance should require more neurons for encoding than a distribution with less variance. Thus, based on Equations 14 and 19 (Appendix) we will introduce a noise threshold criterion to couple the number of neurons to the entropy of the signal distribution to be encoded.

Let the firing rate r^i of each neuron obey the homogenous Poisson model. An input signal s elicits a population response $r = (r^1, \ldots, r^n)$ from which an output signal $u = \max_{\tilde{u}} P(\tilde{u}|r)$ can be retrieved. In the Appendix, Section 'Modelling the response of a neurin population, deconding scheme, and noise' we argue that the noise $R_{\text{noise}} = \{u - F(s);\ u \in [0, 1] \times [0, 1], s \in \mathbb{R}^2\}$, which is generated by the response of the neuron population, can be described by Equation 19. However, the strength of the noise $S_{\text{noise}} = \{F^{-1}(u) - s;\ u \in [0, 1] \times [0, 1], s \in \mathbb{R}^2\}$ which actually acts on the retrieving procedure of the input from the neuronal response depends on the entropy of the distribution of input signals (Equation 14). We assume that the strength of noise $H(S_{\text{noise}})$ is constant for all retinal positions and set $H(S_{\text{noise}}) = T_{\text{noise}}$ with a fixed $T_{\text{noise}} = \text{constant}$. Furthermore, to account for the different scales of retinal speed at different positions of the retina we estimated the entropy in terms of the logarithm of retinal speed rather than from retinal speed itself. Thus, fixing the noise T_{noise} and inserting Equation 19 in Equation 14 the number of neurons which encode for the distribution with a certain entropy $H(\Phi, \log(V))$ can be calculated by

$$N = \exp(-(2.05 + T_{\text{noise}} - H(\Phi, \log(V))/\sqrt{2}). \tag{4}$$

For each distribution, the entropy $H(\Phi, \log(V))$ is estimated by the k-nearest neighbor-method (Kozachenko and Leonenko 1987).

The calculation of the tuning properties of the neurons

From the mean response of the neurons (Equation 3) we can extract separate tuning curves for retinal speed and retinal direction. These tuning curves allow quantitative statements about the preferred speed and direction and about the respective tuning widths. We calculated the tuning curves of a neuron with receptive field center (c_1, c_2) by determining the position of maximal response $(\Phi_{\max}, V_{\max}) = F^{-1}(c_1, c_2)$ in the retinal velocity space while keeping either retinal speed ($V_{\max}$) or retinal direction ($\Phi_{\max}$) fixed and sampling retinal direction and retinal speed, respectively.

From this tuning curves we analyze the mean direction Φ_{mean}, which is the average over all directions weighted by the mean neural response $r_m(\Phi, V_{\max})$, and which is represented by the unit vector

$$\vec{n}(\Phi_{\text{mean}}) := \frac{\int \vec{n}(\Phi) r_m(\Phi, V_{\max}) \mathrm{d}\Phi}{\int r_m(\Phi, V_{\max}) \mathrm{d}\Phi},$$

where $\vec{n}(\phi)$ is the unit vector pointing in the direction ϕ, the directional tuning width

$$\Phi_{\text{width}} := 2\sqrt{\frac{\int |\Phi_{\text{mean}} - \Phi|^2_{\text{circ}} r_m(\Phi, V_{\max}) \mathrm{d}\Phi}{\int r_m(\Phi, V_{\max}) \mathrm{d}\Phi}},$$

the preferred speed $V_{\max}$, and the speed tuning width $V_{\text{width}} := |V_{hr} - V_{hl}|$ defined by the difference between the two speeds V_{hl} and V_{hr} at the half mean response $r_m(\Phi_{\max}, V_{hl}) = r_m(\Phi_{\max}, V_{hr}) = r_m/2$. We take the mean direction Φ_{mean} rather than the direction $\Phi_{\max}$ to characterize the directional tuning properties of the neurons because many neurons, particularly the broadly tuned ones, have nonsymetrical and skew tuning curves where the $\Phi_{\max}$ differs from Φ_{mean}. In this case Φ_{mean} is more informative about the direction encoding properties and the directional preference of the neuron than $\Phi_{\max}$. To characterize the skewness Υ of the tuning curve we use the expression

$$\Upsilon = \begin{cases} 2(\Phi_{\text{mean}} - \Phi_{\max})/\Phi_{\text{width}}; & -\pi \le (\Phi_{\text{mean}} - \Phi_{\max}) \le \pi \\ 2(\Phi_{\text{mean}} - \Phi_{\max} - 2\pi)/\Phi_{\text{width}}; & (\Phi_{\text{mean}} - \Phi_{\max}) > \pi \\ 2(\Phi_{\text{mean}} - \Phi_{\max} + 2\pi)/\Phi_{\text{width}}; & (\Phi_{\text{mean}} - \Phi_{\max}) < -\pi \end{cases},$$

which is a simple version of skewness ((mean – mode) / SD) adapted to the metric on the circle.

Results

Figure 7(a) shows the number of neurons at each position in the visual field for $T_{\text{noise}} = -9$. It can be clearly seen that there is a peak (2809) in the number of neurons which encode for positions close to the fovea. The number of neurons for each position in the visual field strongly drops from foveal positions to far less values (324) for the outmost peripheral positions. As argued above the number of encoding neurons reflects the variability of the distribution which is to be encoded and the resulting neuron numbers are according to the properties of the statistics of retinal

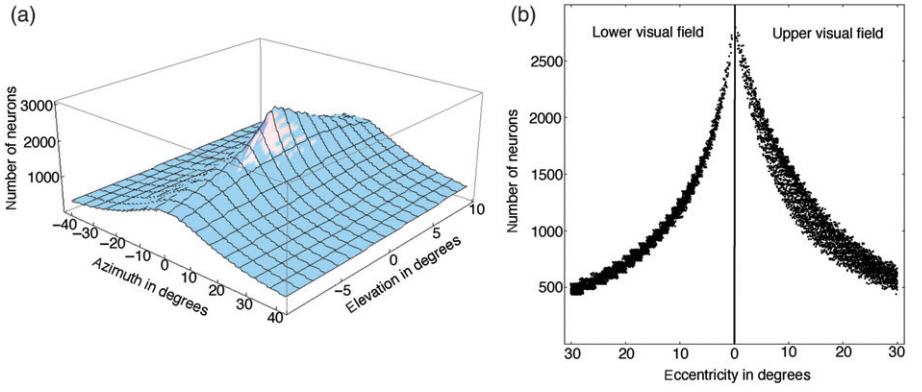

Figure 7. (a) Numbers of neurons encoding retinal velocity for each position of the visual field, such that allowed noise entropy is $H(S_{\text{noise}}) = -9$ for each position. The peripherally decreasing entropy of the distribution of motion signals leads to peripherally decreasing neuron numbers. Changing the allowed noise entropy change the absolute number of neurons but keeps the ratios between the neuron numbers of different positions constant. (b) Numbers of neurons for each position in the visual field plotted over the eccentricity for the lower visual field (right) and the upper visual field (left).

motion presented in (Calow and Lappe 2007). There are also differences in the numbers of neurons for the populations of the upper and the lower visual field. Since variability is higher for the upper visual field, neuron numbers faster decrease with eccentricity for the lower visual field (Figure 7b).

These characteristics of neuron numbers over the visual field can be regarded as a magnification characteristics linked to the local statistics of retinal motion. However, the variability of motion signals is only one aspect that may shape the magnification characteristics. The magnification found in electro-physiological recordings (Albright and Desimone 1987) might reflect multiple aspects of behavior in natural environments. Thus, magnification, which is generally high for foveal positions and rather weak for peripheral positions, and which is regarded to be directly linked to the spatial acuity of perception, can be found in all areas of the visual cortex independent of motion sensitivity. Therefore, it is difficult to compare the magnification characteristics directly to the local statistics of retinal motion signals.

In the following, we will describe the tuning properties of the effectively encoding neuron populations. The properties result from the sensitivity of the neurons for retinal velocity which can be calculated by the mapping of the circular gaussian receptive fields onto the velocity space in terms of the transformation function F (Equation 3). Thus, based on the statistics of optic flow the concept of efficient encoding by mutual information maximization allows to predict receptive fields of efficiently encoding neurons in the retinal velocity space for each position of the visual field (Figure 6b).

Figure 8 shows the direction tuning curves of the neurons of Figure 6(b). These tuning curves derived from the statistics of optic flow illustrate the variety of shapes and peculiarities found in the distribution of direction tuning curves. They are a direct consequence of the efficient encoding scheme and rely on the statistics of optic flow and the respective transformation function. The examples demonstrate

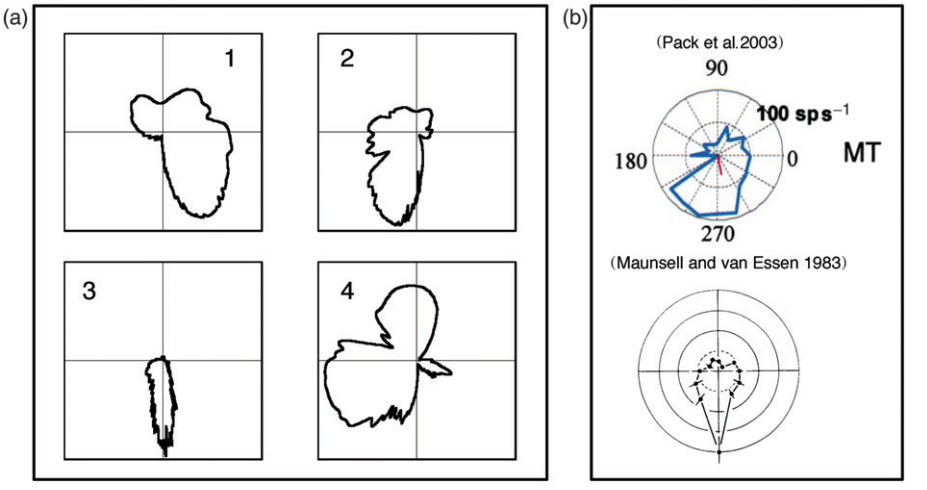

Figure 8. (a) Directional tuning curves of the four example neurons from Figure 6 showing broad (1, 2, 4) or sharp tuning (3). The specific shapes of the tuning curves are caused by the respective position of their receptive fields in the direction space. (b) Example directional tuning curves of neurons in area MT (Maunsell and van Essen 1983; Pack et al. 2003).

the variety of tuning properties that may be found at a single visual field location. Some neurons can be characterized as sharply tuned (Figure 8a3), others are broadly tuned with a skewed tuning curve (Figure 8a1 and 2). Still others are sensitive in more than one direction (Figure 8a4). Although these tuning curves are examples obtained for one particular position in the visual field, similar shapes of tuning curves are found for all positions in the visual field even though the proportions between sharply and broadly tuned curves differ. The spectrum of tuning curves resembles directional tuning curves of neurons in area MT of the macaque monkey (see Figure 8b).

Figure 9(a) shows the speed tuning curves of the neurons of Figure 6(b). The speed tuning curves appears to resemble speed tuning curves measured in monkey area MT (Figure 9b). Like speed tuning in MT (Lagae et al. 1993) the speed tuning curves derived from the statistics of optic flow can be classified as low pass (Figure 9a1), tuned (Figure 9a1 and 2), and high pass (Figure 9a4).

After describing these example neurons, we next consider the distribution of properties over the neuron population. Figure 10 shows distributions of preferred stimuli (best responses (Φ_{max}, V_{max})) for the neuron populations at some example positions in the visual field. Each dot in each panel marks the selectivity of a single neuron in the retinal velocity space. Clearly, the distribution of sensitivity differs for different positions in the visual field and reflects the statistics of motion signals at that visual field position (Figure 4).

As a first way to quantify the position dependence of the neuron properties we analyze the dependence of the tuning properties on the eccentricity of the receptive field center. Thus, we group together all neurons lying close to the fovea within an eccentricity of 2° (parafoveal representation) and all neurons lying between 2° and 30° (peripheral representation). In a second examination, we compare the tuning properties of the neuron populations for the upper and the lower visual field.

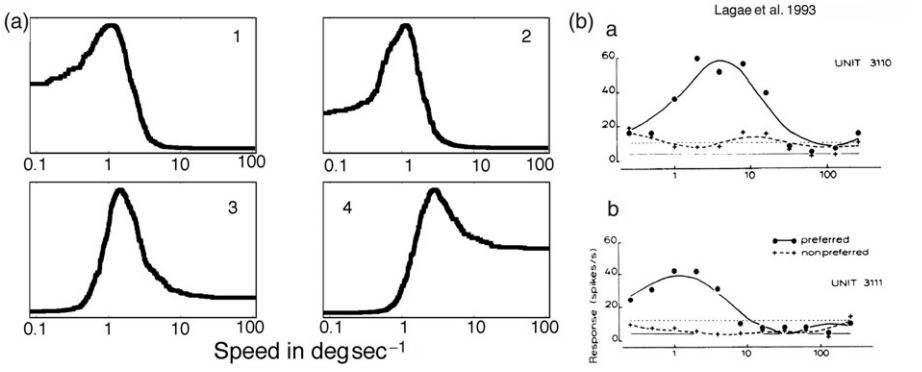

Figure 9. (a) Speed tuning curves of the four example neurons from Figure 6 showing low pass (1), tuned (2, 3), and high pass (4) characteristics. (b) Example speed tuning curves of neurons in area MT (Lagae et al. 1993).

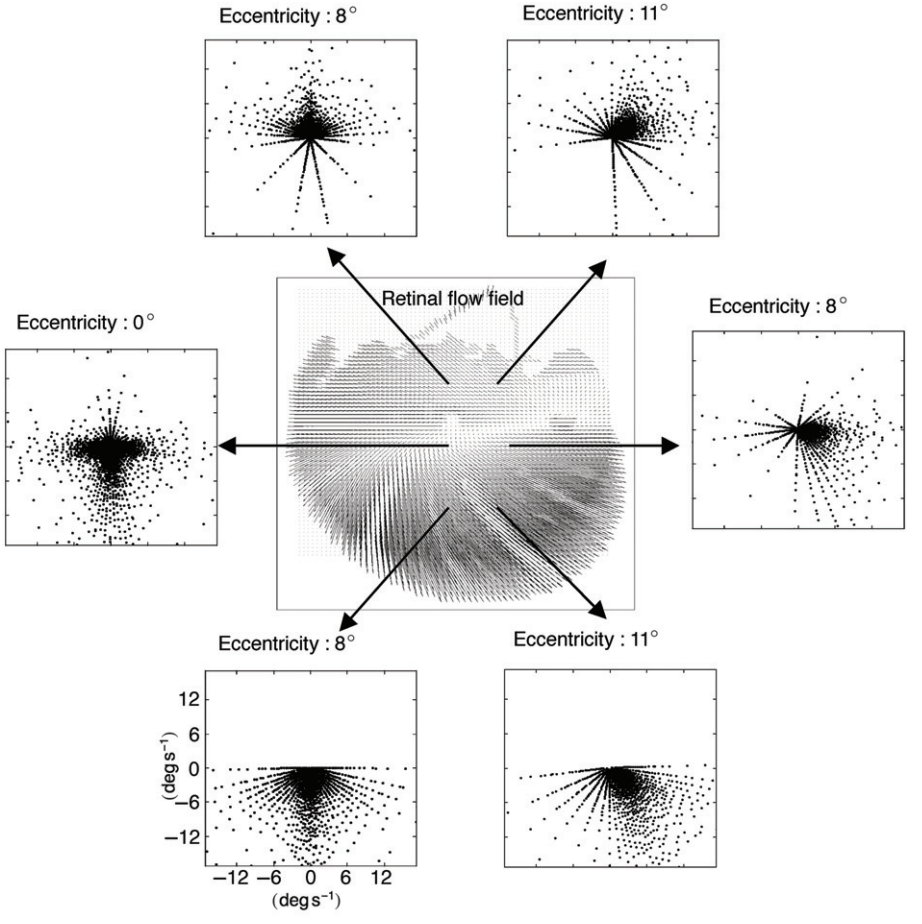

Figure 10. Examples of velocity preference distributions at several visual field locations. Dots in each panel show preferred 2D velocities for single neurons from the distributions.

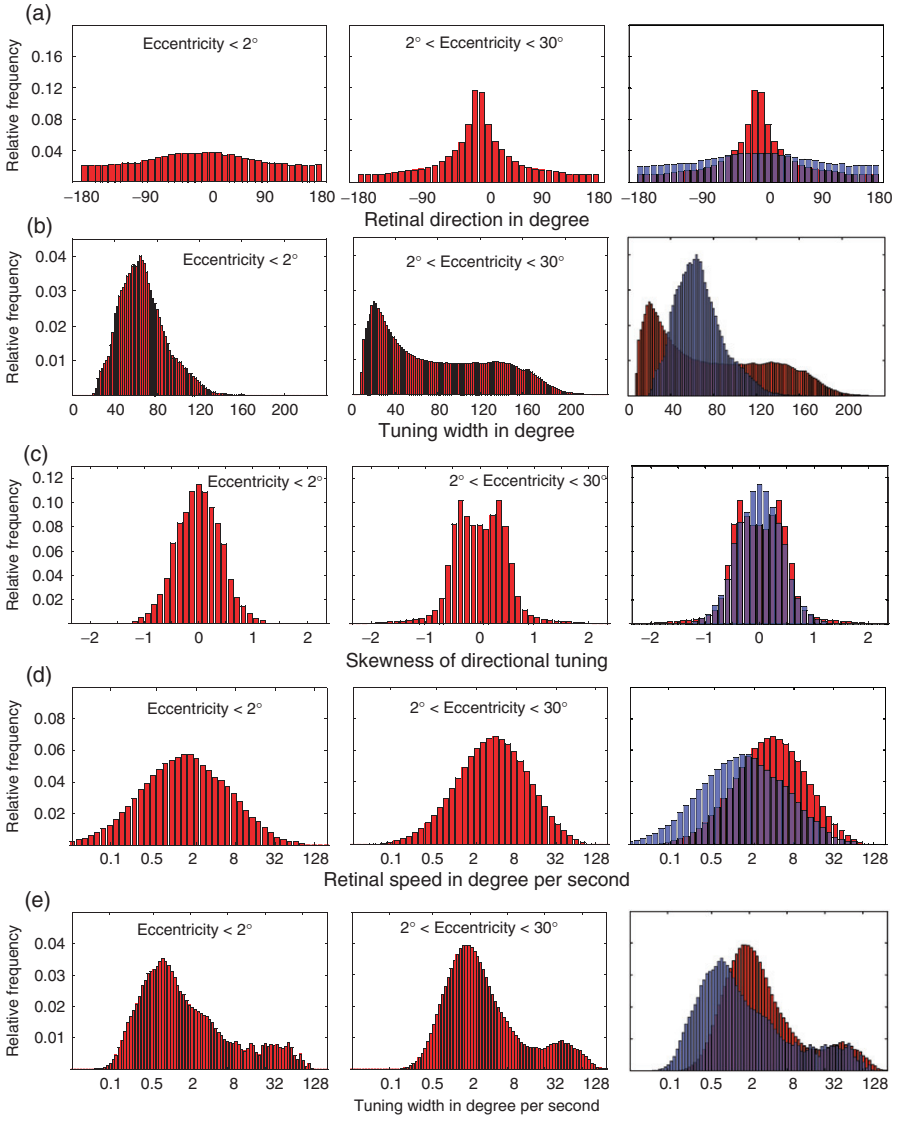

Figure 11. Tuning properties of the model neuron populations within eccentricities of 2° and between 2 and 30°, respectively, and for both classes superimposed for the sake of comparability. (a) Distributions of preferred directions plotted as deviations from the radial direction, (b) Distributions of directional tuning widths, (c) Distribution of directional skewness, (d) Distributions of preferred retinal speeds, (e) Distributions of retinal speed tuning widths.

Figure 11(a) depicts the distribution histograms of direction preferences Φ_{mean} for the parafoveal neuron population within eccentricities of 2° and for the peripheral neural population with eccentricities between 2° and 30°. The bin width for these plots is 10°. The distributions of preferred direction show a bias for the radial direction ($\Phi_{\text{mean}} = 0$). This bias is rather weak for the parafoveal positions but is strongly pronounced for peripheral positions. This relates to the statistics of

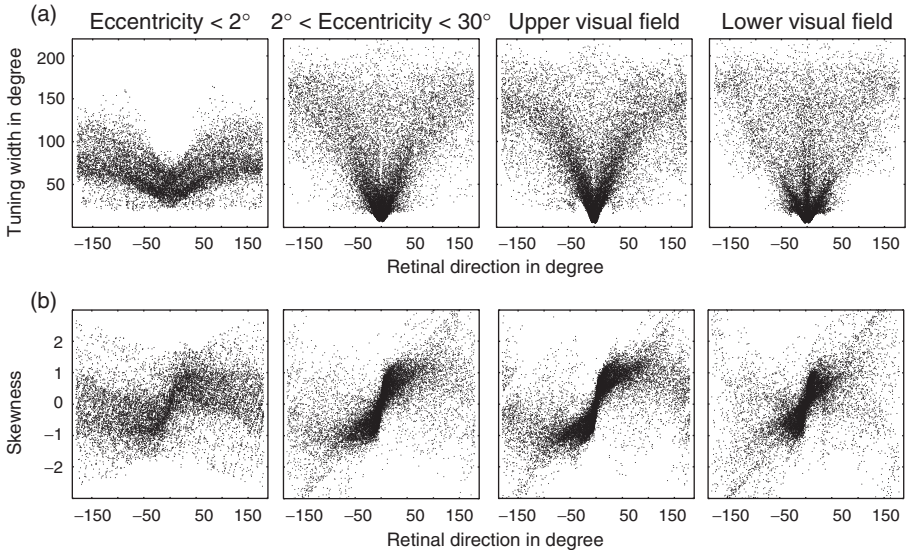

Figure 12. (a) Directional tuning width of neurons plotted over their according preferred tuning direction. (b) Skewness of directional tuning of neurons plotted over their according preferred tuning direction, The panels show the tuning properties of parafoveal and peripheral neurons and for neurons of the upper and the lower visual field.

retinal motion signals which are increasingly radial at higher eccentricities (Calow and Lappe 2007). Figure 11(b) shows distributions of the width of directional tuning. The bin width for these plots is 2°. The parafoveal neuron population shows a peak around 63° and covers a range between 19° and 140° tuning width. The peripheral population shows the peak of the distribution of tuning widths at a smaller value around 23° but the distribution becomes broader, and covers a range between 7° and 200° tuning width. The reason for this is that for higher eccentricities the neuronal sensitivity for directions close to radial are more narrowly tuned, and neurons sensitive for directions away from radial or opposite of radial are broadly tuned. For the parafoveal neuron populations tuning is more uniform between neurons and overall sharper.

This connection between tuning width and preferred direction for the parafoveal and for the peripheral neuron populations is illustrated in Figure 12(a). Figure 11(c) depicts the distributions of the skewness of directional tuning. For the parafoveal neuron population (eccentricity smaller than 2°), the skewness of the tuning curves is symmetrically distributed around zero and the distribution covers a range between −1 and 1. The peripheral neuron population shows two skewness peaks around −0.353 and 0.353. While the proportion of non-skewed tuning curves is high the distribution of directional skewness has long tails and covers a range which significantly exceeds −1 and 1. The skewness properties reflect the statistics of retinal motion signals (Calow and Lappe 2007). Encoding mostly nonskewed distributions of retinal motion directions, which occur particularly for eccentricities smaller than 2°, will result in tuning properties whose skewness is symmetrically distributed and has a maximum at zero. To explain this issue imagine you have a nonskewed distribution to encode. Following the encoding scheme proposed above the encoding is done by uniformly covering the image of the cumulative

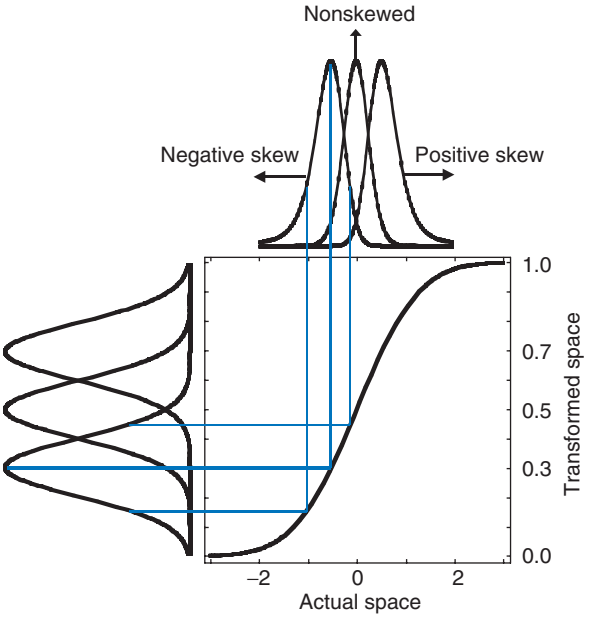

Figure 13. Mapping of symmetric tuning curves onto the actual space in terms of an example transformation function F, which is the cumulative distribution function of a Gaussian with SD $\sigma = 1$. The resulting tuning curves are systematically skewed. The blue lines illustrate the transformation rule.

distribution function with gaussian tuning curves. These tuning curves are skewed in the original space (see Figure 13). Since, the highest density of tuning curves is around the position of the peak of the original distribution, tuning curves around zero skewing form the highest proportion of the population and the skewing properties are symmetrically distributed around zero. For skewed distributions of retinal motion directions, which appear for peripheral positions in the visual field, the maximum of tuning skewness shifts to nonzero values. This is evident by the same argument namely that the highest density of tuning curves is around the position of the peak of the original distribution. However, these tuning curves are no longer nonskewed. The peaks for the distribution of directional tuning skewness for the peripheral neuron populations reflect the two different signs of skewness of the distribution of retinal directions for the different quadrants of the visual field (Calow and Lappe 2007). Concerning the relation between preferred tuning direction and skewness, Figure 12(b) shows that the connection between preferred direction and skewness is rather weak for foveal neuron populations but strong for the peripheral neuron population, for which most neurons have directional tuning curves skewed toward the radial direction.

Figure 11(d) depicts the distributions of preferred speed on the logarithmic scale for the parafoveal and the peripheral neuron populations. Bin width is 0.2 in logarithmic units, which is calculated by

$$\text{bin width} = \frac{\log\left(512\,\frac{\text{deg}}{\text{s}}\right) - \log\left(0.01\,\frac{\text{deg}}{\text{s}}\right)}{\text{number of bins}} = \frac{\log\left(\frac{512}{0.01}\right)}{\text{number of bins}}.$$

The distribution of preferred retinal speeds is clearly shifted to higher velocities for the peripheral neuron population. The peak of the preferred speed distribution is around $1.3^\circ\,s^{-1}$ for the parafoveal and around $3.1^\circ\,s^{-1}$ for peripheral neuron population. On the logarithmic scale, the distribution for the parafoveal positions is slightly broader than for the peripheral positions. For the parafoveal neuron population, 98% of neurons have preferred speeds between $0.03^\circ\,s^{-1}$ and $22.1^\circ\,s^{-1}$. For the peripheral neuron population 98% of neurons have preferred speeds between $0.12^\circ\,s^{-1}$ and $34.04^\circ\,s^{-1}$. However, one must be careful to note that this does not mean that no neurons are sensitive for higher retinal speed. Neurons with high pass characteristics respond also to much larger speed values. Figure 11e shows the distributions of speed tuning widths on the logarithmic scale. Bin width is 0.1. Tuning curves are overall broader at peripheral positions. The distribution of tuning widths for parafoveal positions covers a large range starting with $0.1^\circ\,s^{-1}$ and ending around $95^\circ\,s^{-1}$. It has a peak around $0.65^\circ\,s^{-1}$. Neurons of the peripheral representation cover a range of tuning widths between $0.27^\circ\,s^{-1}$ and $132^\circ\,s^{-1}$ with a maximum at $1.54^\circ\,s^{-1}$. Furthermore, the distributions of speed tuning widths have an unique shape. 90% of the neurons have tuning widths not larger than 26 and $36^\circ\,s^{-1}$ for parafoveal and peripheral positions, respectively. But the remaining 10% of neuron form a second bulge of the distribution. This bulge with extremely large tuning widths is more pronounced for distribution encoding higher eccentricities. The properties of the distribution of preferred retinal speed and speed tuning width over the visual field therefore reflect the findings in (Calow and Lappe 2007), where it is shown that retinal speed increases for more peripheral positions particularly for natural scenes and when gaze-stabilizing eye movings are performed.

Retinal optic flow has different statistical properties in the upper and lower visual fields (Zanker and Zeil 2005; Calow and Lappe 2007). The distribution of retinal flow velocities in the lower visual field compared to the upper visual field is shifted to higher speeds and more biased to radial directions. This difference can also be appreciated from the example distributions of retinal motion signals in Figure 4. The different statistical properties of retinal flow in the upper and lower visual fields are caused by differences in the statistics of depth in natural scenes between the upper and the lower visual field (Calow and Lappe 2007). In the efficient encoding scheme, the differences in the statistical properties of retinal flow should be reflected in the tuning properties of the neuron population. We therefore investigated the tuning properties for the neurons separately for the upper and the lower visual field. Figure 14 shows the tuning properties for the neuron populations of the upper visual field (left column), and of the lower visual field (middle column). Differences between the upper and lower visuel field are seen in all five of the analyzed parameters. The radial bias for the directional preferences of the neurons is more pronounced in the lower visual field (Figure 14(a)). Directional tuning is overall sharper in the lower visual field (tuning width peaks at 19° compared to 27°, Figure 14(b)) but also ranges higher (up to the 160 to 200°range) such that also more very broadly tuned neurons exist in the lower visual field. Since the neuronal sensitivity for directions close to radial are more narrowly tuned (because of the radial bias of the directional statistics of motion signals, see also Figure 12a), and neurons sensitive for directions away from radial or opposite of radial are broader tuned, these results reflect again the statistical directional properties of retinal motion signals. Also the distributions of directional skewness of tuning curves

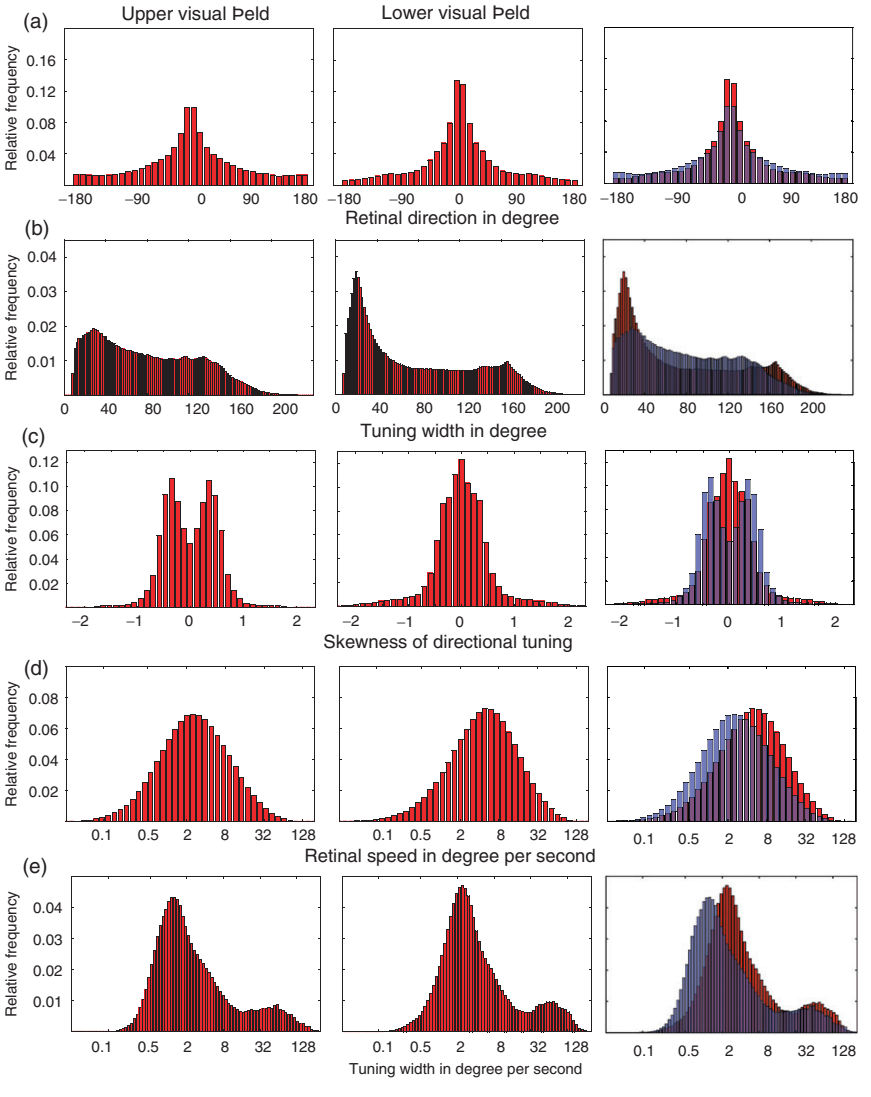

Figure 14. Tuning properties of the model neuron populations for the upper and lower visual fields separately and for both superimposed for the sake of comparability. (a) Distributions of preferred directions plotted as deviations from the radial direction, (b) Distributions of directional tuning widths, (c) Distribution of directional skewness, (d) Distributions of preferred retinal speeds, (e) Distributions of retinal speed tuning widths.

Figure 14(c) show remarkable differences between the upper and the lower visual field. Whereas for the upper visual field the skewness distribution develop two peaks at -0.353 and 0.353 and has a rather deep indent for zero skewness, the skewness distribution for the lower visual field is symmetrically distributed around a peak at zero skewness. These results are due to the findings in Calow and Lappe (2007) that particularly when stabilizing eye movements are involved the distributions of the directions of motion signals develop high values of skewness for large domains of the upper visual field, where skewness is less pronounced for the lower visual field.

Preferred speeds are on average higher in the lower visual field (peak at $4^\circ\,s^{-1}$ compared to $2.2^\circ\,s^{-1}$, Figure 14(d)). The proportion of broadly speed tuned neurons is higher in the lower visual field (Figure 14(e)). Thus, the properties of the neurons encoding for the lower visual field are adapted to higher retinal speed and more uniform direction, whereas the neurons encoding for the upper visual field are better suited for lower retinal speed in accordance to the statistics of retinal flow. The differences in the properties of neuron populations between the upper and the lower visual field might be actually stronger than shown here. For technical reasons the visual field in our investigations was restricted to only 20° in elevation, whereas the actual visual fields of humans and other primates are much larger in the vertical range.

Discussion

We presented a method to construct an efficient encoding scheme directly from a prior distribution of input stimuli. This method is based on maximizing the mutual information between input and output in the presence of internal noise and under the assumption that the range of the output is limited. This efficient encoding scheme leads to a population of processing units or neurons for which the population response transforms the input distribution into a uniform distribution. Thus, in the output all redundancies in the statistics of the input signals are dissolved, each processing unit handle the same load of information, and the internal noise is optimally dispersed over the input space. Thus, considering the possible range of input signals and their statistics the output is as tightly as possible linked to the input.

We applied this efficient encoding scheme to the realm of motion processing to find possible connections between the properties of the motion processing pathway of the brain and the statistics of retinal flow induced by self motion during the natural environment. We identified analogies between the properties of efficient encoding neurons and electro-physiological findings, and predict possible properties of the velocity tuning in the motion processing pathway of the brain.

We acquired prior distributions of retinal motion signals based on the method introduced in Calow and Lappe (2007). These are distributions of true retinal velocity calculated from geometrical optics rather than motion extracted by early motion estimators from spatial-temporal image luminance variations. Our procedure makes the results independent of the properties of particular motion detectors and their notorious performance problems (Barron et al. 1994). However, since early motion detection is a necessary part of the motion processing pathway the errors made by in motion detection are modeled by internal noise added to the signal by processing. Nevertheless, we were able to reproduce some of the properties of real motion sensitive neurons in area MT rather directly from the statistics of retinal motion. Similar to the findings in (Lagae et al. 1993), individual neurons in the efficient encoding scheme can be characterized as low-pass, high-pass, or tuned in terms of their speed selectivity. Furthermore, the directional tuning curves of individual neurons resemble those from electro-physiological recordings (Figure 8). The emergence of different categories of tuning curves (sharply tuned and broadly tuned for direction; low pass, tuned, and high pass for speed) in the

efficient encoding model is not as arbitrary as it may appear. It is a consequence of the bending of the actual input space into a bounded output range along with the efficient encoding procedure. Thus, neurons with receptive fields close to the bounds of the output space have extended receptive fields in the input space, whereas neurons with receptive fields close to the center of the output space have very narrow tuning curves in the input space (compare Figure 6a with Figures 8a and 9a). Therefore, the type of the tuning curve is correlated with the preferred speed and direction of the respective neuron. This correlation may be investigated in future experimental studies . Considering the tuning properties of the entire neuron population we propose that they completely reflect the statistical properties of retinal motion. Consequently, the neuron population shows a radial bias of the directional tuning which is increased for higher eccentricities. Moreover, preferred retinal speeds shift to higher values for higher eccentricities.

A radial bias of the preferred directions of motion sensitive neurons in area MT of the macaque monkey was reported by (Albright 1989) following earlier reports in the cat (Rauschecker et al. 1987). However, the bias in area MT is weaker than the one found in the present study and weaker than the bias in cat area PMLS. In monkey area MT, a significant radial bias is revealed only in the representation of the peripheral visual field (positions between 12° and 30° eccentricity in the study of Albright 1989). In our model, a bias exists already for parafoval positions (larger than 2° eccentricity), but, consistent with findings from area MT, vanishes in central vision. Aside from the fact that the number of neurons considered in our study (around 10^7) largely exceeds the number of recorded neurons in (Albright 1989) (438), the differences in the parafoveal region might be explained by observing that area MT is not exclusively devoted to optic flow processing but is also involved in other visual motion tasks such as, for instance, smooth pursuit eye movements and eye-hand coordination. For instance, observing the motion of the hands while being engaged in manual tasks could provide a statistically considerable amount of retinal motion toward the fovea, which is not taken into account in this study.

The distributions of preferred speeds in area MT is shifted to somewhat higher speeds (Maunsell and van Essen 1983) than the speed distributions extracted in the present study. Whereas (Maunsell and van Essen 1983) state a peak at $32^\circ\,s^{-1}$ for 89 recorded neurons, in the present study the peak is identified at $3.1^\circ\,s^{-1}$. This might be explained by observing that the present investigation is based on a model of human locomotion behavior. Since macaque monkeys are much smaller than humans they would experience higher visual velocities when moving on the ground. Thus, adaptations of the motion processing slightly differ. On the other hand, since because of technical reasons we have no optic flow data up to 10° downwards in the visual field and therefore we do not take into account neuron populations which encode for these positions, we should acknowledge that leaving aside this neuron population may cause an underestimation of the mean preferred speed of the encoding neurons within 30° eccentricity.

Receptive fields of neurons in area MT of the macaque monkey are several degrees in diameter and increase in size for peripheral neuron populations (Albright and Desimone 1987). The tuning parameters of our model neuron, in contrast, are based on the statistical properties of the retinal flow at points in the visual field. With respect to first-order statistical properties the tuning for

a particular visual field position can be extended to a larger, peripherally increasing area. Such a receptive field implements a space-variant filter that increases the robustness of the flow representation against noise (Calow et al. 2005). However, second-order statistical properties of the optic flow, like joint motion statistics of neighboring positions within an extended receptive field, are also very interesting and likely to contribute to optic flow analysis. Preliminary investigations suggest that the size of receptive fields in area MT might be correlated to the strong statistical correlations between the motion statistics within these receptive field areas. We expect that the investigation of second-order statistics within yet larger receptive fields (like in MST) might reveal more detailed receptive field substructures that are beneficial for figure ground segregation (Orban et al. 1992, 1995), 3D orientation of planar image patches (Xiao et al. 1997), or heading estimation (Beintema et al. 2004).

The analysis of the properties of the efficient encoding neurons derived from the acquired statistics of retinal motion signals provides further predictions of tuning properties in the neural population. For instance, the results on the directional and speed tuning widths and on the skewness of directional tuning predict connections between tuning width and preferred direction that have not yet been investigated in physiological data. Also the differences between the tuning properties of the efficiently encoding neuron populations of the upper and the lower visual field serve as predictions, which can be tested with electro-physiological recordings in monkeys. Our results also predict different neuronal magnification characteristics in the retinal velocity space (Figure 10) for different positions in the visual field. At a given position in the retinotopic map neuronal motion sensitivities should be clustered around those with high occurrence probability in the retinal flow and neuron density in the tails of retinal motion distribution should be lower. This might be tested by measuring the motion sensitivities of several neurons where receptive field centers are around a certain position in the visual field. Furthermore, our results show that along with the smaller variability of motion signals for more peripheral positions the number of neurons can peripherally decrease while preserving the level of noise. Thus, supposing that this magnification is superimposed on the magnification due to spatial acuity our results predict that magnification will be different in brain areas which are motion sensitive and which are not motion sensitive.

Our study is modeled on human locomotion. Since locomotion properties differ for other species, the results may not be one-to-one compared to data from other species. Nevertheless the results reflect many aspects of motion processing obtained by electro-pyhsiological recordings from experimentally accessible animals and thus we assume that also other predictions emanating from the efficient encoding scheme applied to human-like motion situations can at least qualitatively serve as predictions for other species. For example, it is reasonable to assume that differences between the encoding properties of the neurons of the upper and the lower visual field should also be found in other primates or even in cats, because the main reason for these differences is the overall statistics of natural scenes and the statistical differences between the ground and the scene over ground (Zanker and Zeil 2005; Calow and Lappe 2007).

Several interesting issues remain for future work. For example, the Bayesian decoding scheme may be applied to develop computational models that detect

retinal motion in the statistically likely range. Secondly, since the constructed neuron populations imply certain motion detection characteristics it may be tested psychophysically whether humans show a similar characteristics in motion detection. The scenes in our study comprises urban and forest scenes. They can be regarded as typical scenes that humans encounter in normal life. Other animal species which live in very different environments (e.g., treeless deserts or rocky gorges) would encounter different flow fields. Thus, an interesting question is how the environment affects the properties of the motion processing pathway. Although we argue that there is evidence that neurons in MT/V5 are mostly sensitive for true retinal motion as a result of space variant filtering we acknowledge that contrast statistics and their influence on motion detection and efficient encoding should be studied and investigated further on. Such examinations would require different input material like real camera movies, which imitate human locomotion. Interesting methods to acquire such data are proposed in (Betsch et al. 2004, Zanker and Zeil 2005, Einhäuser et al. 2007).

As mentioned above we do not take into account all aspects of motion processing. Our focusing on the first order or local statistics of optic flow prevents us from statements about the inner structure of motion sensitive receptive fields. Furthermore, receptive field sizes and magnification characteristics might be evolutionary optimized according to higher-order statistics of optic flow. Therefore, future work will focus on issues which involve statistical linkages between motion of different positions of the visual field and the encoding schemes for these problems. First results concerning higher-order statistics of optic flow can be found in (Roth and Black 2007). In this study among other things principal components of optic flow due to 5×5 image patches for vertical and horizontal image motion are found. Future studies will go beyond this issues by searching for the independent components of retinal motion within spatial domains depending on the position in the visual field.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Albright TD. 1989. Centrifugal directionality bias in the middle temporal visual area MT of the macaque. Visual Neuroscience 2:177–188.

Albright TD, Desimone R. 1987. Local precision of visuotopic organization in the middle temporal area MT of the macaque. Experimental Brain Research 65:582–592.

Atick JJ. 1992. Could information theory provide an ecological theory of sensory processing? Network: Computation in Neural Systems 3(2):213–251.

Barlow HB. 1961. Possible principles underlying the transformation of sensory messages. In: Rosenblith WA, editor. Sensory Communication. Cambridge, MA: MIT Press. pp 217–234.

Barron JE, Fleet DJ, Beauchemin SS. 1994. Performance of optical flow techniques. International Journal of Computer Vision 12:43–77.

Beintema JA, van den Berg AV, Lappe M. 2004. Circular receptive field structures for flow analysis and heading detection Dordrecht, Boston, London: Kluwer Academic Publishers. pp 223–248.

Betsch BY, Einhäuser W, Körding KP, König P. 2004. The world from a cat's perspective – statistics of natural videos. Biological Cybernatics 90:41–50.

Calow D, Krüger N, Wörgötter F, Lappe M. 2004. Statistics of optic flow for self-motion through natural sceneries. In: Ilg U, Bülthoff HH, Mallot HA, editors. Dynamic perception, workshop of the gi section 'computer vision'. Berlin: Akademische Verlagsgesellschaft Aka GmbH. pp 133–138.

Calow D, Krüger N, Wörgötter F, Lappe M. 2005. Biologically motivated space-variant filtering for robust optic flow processing. Network 16(4):323–340.

Calow D, Lappe M. 2007. Local statistics of retinal optic flow for self-motion through natural sceneries. Network 18(4):343–374.

Duffy CJ, Wurtz RH. 1991. Sensitivity of MST neurons to optic flow stimuli. i. a continuum of response selectivity to large-field stimuli. Journal of Neurophysiological 65:1329–1345.

Einhäuser W, Schuhmann F, Bardins S, Bartl K, Böning G, Schneider E, König P. 2007. Human eye-head co-ordination in natural exploration. Network: Computation in Neural Systems 18(3):267–297.

Elder EH, Goldberg RG. 2002. Ecological statistics of Gestalt laws for the perceptual organization of contours. Journal of Vision 2:324–353.

Huang J. 2000. Statistics of natural images and models. Thesis, Providence Rhode Island: Applied Mathmatics at Brown University. pp 1–129.

Imai T, Moore ST, Raphan T, Cohen B. 2001. Interaction of the body, head, and eyes during walking and turning. Experimental Brain Research 136(1):1–18.

Kozachenko LF, Leonenko NN. 1987. Sample estimate of the entropy of a random vector. Problemy Peredachi Informatsíe 23:95–101.

Krüger N, Wörgötter F. 2002. Multi-modal estimation of collinearity and parallelism in natural image sequences. Network Computation in Neural Systems 13:553–576.

Lagae L, Raiguel S, Orban GA. 1993. Speed and direction selectivity of macaque middle temporal neurons. Journal of Neurophysiology 69(1):19–39.

Lappe M. 2000. Computational mechanisms for optic flow analysis in primate cortex. In: Lappe M, editor. Neuronal processing of optic flow, volume 44 of International Review & Neurobiology. San Diego, London: Academic Press. pp 235–268.

Lappe M, Bremmer F, Pekel M, Thiele A, Hoffmann KP. 1996. Optic flow processing in monkey STS: A theoretical and experimental approach. Journal of Neuroscience 16(19):6265–6285.

Lappe M, Bremmer F, van den Berg A.V. 1999. Perception of self-motion from visual flow. Trends in Cognitive Sciences 3(9):329–336.

Lappe M, Hoffmann K-P. 2000. Optic flow and eye movements. In: Lappe M, editor. International Reviev of Neurobiology, volume 44 Neuronal Processing of Optic Flow. New York: Academic Press. pp 29–47.

Lappe M, Pekel M, Hoffmann K-P. 1997. Optokinetic eye movements elicited by radial optic flow in macaque monkeys. Abstract – Society for Neuroscience 23:758.

Lappe M, Pekel M, Hoffmann K-P. 1998. Optokinetic eye movements elicited by radial optic flow in the macaque monkey. Journal of Neurophysiology 79(3):1461–1480.

Laughlin SB. 1981. Simple coding procedure enhances a neuron's information capacity. Zeitschrift Fur Naturforschung 36C:910–912.

Linsker R. 1988. Self-organization in a perceptual network. IEEE Computer 88:105–117.

Maunsell JHR, van Essen DC. 1983. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. Journal of Neurophysiology 49(5):1127–1147.

Nadal JP, Parga N. 1994. Nonlinear neurons in the low-noise limit: A factorial code maximizes information transfer. Network: Computation in Neural Systems 5(4):565–581.

Niemann T, Lappe M, Büscher A, Hoffmann K-P. 1999. Ocular responses to radial optic flow and single accelerated targets in humans. Vision Research 39(7):1359–1371.

Orban GA, Lagae L, Raiguel S, Xiao D, Maes H. 1995. The speed tuning of medial superior temporal MST cell responses to optic-flow components. Perception 24:269–285.

Orban GA, Lagae L, Verri A, Raiguel S, Xiao D, Maes H, Torre V. 1992. First-order analysis of optical flow in monkey brain. Proceedings of the National Academy of Sciences of the USA 89:2595–2599.

Pack CC, Livingstone MS, Duffy KR, Born RT. 2003. End-stopping and the aperture problem: Two-dimensional motion signals in macaque V1. Neuron 39:671–680.

Perrone JA, Thiele A. 2001. Speed skills: Measuring the visual speed analyzing properties of primate MT neurons. Nature Neuroscience 4(5):519–532.

Rauschecker JP, von Grünau MW, Poulin C. 1987. Centrifugal organization of direction preferences in the cat's lateral suprasylvian visual cortex and its relation to flow field processing. Journal of Neuroscience 7(4):943–958.

Rodman HR, Albright TD. 1987. Coding of visual stimulus velocity in area MT of the macaque. Vision Research 27(12):2035–2048.

Roth S, Black MJ. 2007. On the spatial statistics of optical flow. International Journal of Computer Vision 74(1):33–50.

Saito H-A, Yukie M, Tanaka K, Hikosaka K, Fukada Y, Iwai E. 1986. Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. Journal of Neuroscience 6(1):145–157.

Solomon D, Cohen B. 1992. Stabilization of gaze during circular locomotion in darkness: II. contribution of velocity storage to compensatory head and eye nystagmus in the running monkey. Journal of Neurophysiology 67(5):1158–1170.

Xiao D-K, Marcar VL, Raiguel SE, Orban GA. 1997. Selectivity of macaque mt/v5 neurons for surface orientation in depth specified by motion. European Journal of Neuroscience 9(5):956–964.

Zanker JM, Zeil J. 2005. Movement-induced motion signal distributions in outdoor scenes. Network 16(4):357–376.

Appendix

Maximizing mutual information: Derivation of the transformation function

Here we derive the transformation function which maximizes the mutual information between output and input for a limited range of the output. This section of the appendix is a general supplement to Section 'Maximizing mutual information' and is not essential to understand the efficient encoding scheme that the present study relies on. It is included since, it lists all assumptions we make about the properties of the channel and the internal noise and because it characterizes the class of nonlinear transformation functions that maximize the entropy of the output and the transfer of information through a channel with bounded output range in the two-dimensional case (see also Nadal and Parga 1994).

Let S and U be the set of input and output signals, respectively. In the following we will restrict our consideration to $S = \mathbb{R}^2$; $s = (s^1, s^2)$. We assume that the output is a two-component vector, where each component is bounded. Thus, up to two scale factors U can be set to $U \subset [0, 1] \times [0, 1]$. We also assume that the channel occupies the entire $[0, 1] \times [0, 1]$-space for decoding such that $U = [0, 1] \times [0, 1]$. Let $P_S(s)$ be the PDF governing the statistics of occurrence of the input signals, $s \in S$, let $P_U(u)$ be the PDF governing the distribution of output signals $u \in U$, and $P(s, u) = P(u|s)P_S(s) = P(s|u)P_U(u)$ the joint PDF. The mutual information $mI(U, S)$ between output and input is defined as

$$mI(U, S) := \int_{\mathbb{R}^2} \int_{[0,1]\times[0,1]} P(s, u) \log_2 \left(\frac{P(s, u)}{P_S(s)P_U(u)} \right) \mathrm{d}^2 s \mathrm{d}^2 u. \tag{5}$$

In the following, we assume that input and output is linked by $u = F(s) + r_{\text{noise}}$, where $F : \mathbb{R}^2 \to [0, 1] \times [0, 1]$; $F(s) = (f^1(s^1, s^2), f^2(s^1, s^2))$ is the bijective transformation function of the channel. The random variable $r_{\text{noise}} = u - F(s)$ denotes the added noise which is postulated to be independent, such that $P(s, u)) = P(u|s)P(s) = Q(u - F(s))P(s)$. Further suppose that F is differentiable. These assumptions lead us to a condition for the transformation function F which maximizes the mutual information between input and output.

Equation 5 can be transformed to

$$\begin{aligned} mI(U,S) &= -\int P_U(u)\log_2(P_U(u))\mathrm{d}^2u + \int Q(u-F(s))P_S(s)\log_2(Q(u-F(s)))\mathrm{d}^2s\mathrm{d}^2u. \\ &= -\int P_U(u)\log_2(P_U(u))\mathrm{d}^2u + \int Q(r_{\text{noise}})\log_2(Q(r_{\text{noise}}))\mathrm{d}^2r_{\text{noise}} \\ &= H(U) - H(R_{\text{noise}}), \end{aligned} \tag{6}$$

where $H(U)$ and $H(R_{\text{noise}})$ are the entropies of the output U and the noise, respectively. Equation 6 shows that the noise entropy is independent from the selection of the transformation function. Thus maximizing the mutual information means searching for a transformation function F which maximizes $H(U)$. Recall that the transformation function F is supposed to be bijective and differentiable. Thus, there exists a unique inverse mapping F^{-1}. Let $\tilde{S} = \{\tilde{s} = F^{-1}(u);\, u \in S_0\}$ be the random variable governed by the PDF $P_{\tilde{S}}(\tilde{s})$. Note that the random variables S and $\tilde{S}$ are from the same space $\mathbb{R}^2$ but their PDFs are in general different due to the noise R_{noise} and $\tilde{s} = F^{-1}(F(s) + r_{\text{noise}})$. Let the expression & denote the wedge product between differential forms. The PDFs $P_{\tilde{S}}(\tilde{s})$ and $P_U(u)$ are mathematically linked by

$$P_{\tilde{S}}(\tilde{s})\mathrm{d}\tilde{s}^1 \wedge \mathrm{d}\tilde{s}^2 = P_U(F(\tilde{s}))\mathrm{d}f^1(\tilde{s}) \wedge \mathrm{d}f^2(\tilde{s}). \tag{7}$$

Equation 7 leads directly to

$$P_{\tilde{S}}(\tilde{s}) = P_U(F(\tilde{s}))|\det(\mathcal{J}_F(\tilde{s}))|, \tag{8}$$

where $\det(\mathcal{J}_F(\tilde{s})) = (\mathrm{d}f^1/\mathrm{d}\tilde{s}^1)(\mathrm{d}f^2/\mathrm{d}\tilde{s}^2) - (\mathrm{d}f^2/\mathrm{d}\tilde{s}^1)(\mathrm{d}f^1/\mathrm{d}\tilde{s}^2)$ is the determinant of the Jacobian of the transformation F. Replacing P_U in Equation (6) by Equation (8) yields

$$H(U) = -\int P_{\tilde{S}}(\tilde{s})\log_2\left(\frac{P_{\tilde{S}}(\tilde{s})}{|\det(\mathcal{J}_F(\tilde{s}))|}\right)\mathrm{d}^2\tilde{s}. \tag{9}$$

From Equations 7 and 8 it is clear that $\int_{\mathbb{R}^2} |\det(\mathcal{J}_F(\tilde{s}))|\mathrm{d}^2\tilde{s} = \int_0^1\int_0^1 \mathrm{d}u^1\mathrm{d}u^2 = 1$. Thus, $|\det(\mathcal{J}_F(\tilde{s}))|$ is itself a PDF and Equation 9 is the negative of a Kullback–Leibler difference. Thus, $H(U)$ will be maximal if and only if

$$|\det(\mathcal{J}_F(\tilde{s}))| = P_{\tilde{S}}(\tilde{s}). \tag{10}$$

Note that $P_{\tilde{S}}(\tilde{s}) = \int_{\mathbb{R}^2} P(\tilde{s}|s)P_S(s)\mathrm{d}^2s$. In the following, we assume that the noise is weak, which means that $P(\tilde{s}|s)$ can be approximated by the delta-function $P(\tilde{s}|s) \approx \delta(\tilde{s} - s)$. Considering the noise as weak is an approach proposed by several authors (Laughlin 1981; Atick 1992; Nadal and Parga 1994) to obtain transfer functions which maximize information. We thus assume that the output of the channel varies only slightly around the input, such that the noise is weak and the error imposed on the input signal by transferring through the channel is small. Otherwise, retrieving useful information about the input from the output signal would not be possible. Appendix Section 'Modeling the response of a neuron population, decoding scheme, and noise' shows how the strength of the output noise can be steered by the number of neurons which encode the input.

Thus, we can approximate Equation 10 by

$$|\det(\mathcal{J}_F(s))| \approx P_S(s). \tag{11}$$

Equation 11 is a generalization of the derivation of the optimal bounded transfer function in the low noise limit for the one-dimensional case in (Nadal and Parga 1994) (see Equation 16 there) to two-dimensional input. Putting the solution (10) in Equation 8 leads to the conclusion that each transformation function F which transforms the input distribution into a uniform distribution over the image space $(P_U(F(\tilde{s})) = 1, \forall\, \tilde{s};\ F(\tilde{s}) \in [0,1] \times [0,1])$ maximizes the mutual information between input and output. For an example of such a function assume that the PDF $P_S(s) = P(s^2|s^1)P(s^1)$ of the input is known. Let $H_s : \mathbb{R} \to \mathbb{R}$ be the Heaviside step function:

$$H_s(x) = \begin{cases} 0; & x < 0 \\ 1; & x \geq 0. \end{cases}$$

A transformation function F fulfilling (11) can be constructed by

$$f^1(s^1, s^2) = f^1(s^1) = \int_{-\infty}^{\infty} H_s(s^1 - t^1)P(t^1)\mathrm{d}t^1 \tag{12}$$

$$f^2(s^1, s^2) = \int_{-\infty}^{\infty} H_s(s^2 - t^2)P(t^2|s^1)\mathrm{d}t^2. \tag{13}$$

In conclusion, maximizing the mutual information by the transformation function F is accomplished by stretching and compressing the original space such that the noise added by the channel is optimally dispersed over the signal distribution. Thus, signals with a high probability of occurance are less affected by noise whereas signals on the tails of the signal PDF are more affected by noise. Note that the actual noise S_{noise}, which can be retrieved by the back transformation $s_{\text{noise}} = s - F^{-1}(F(s) + r_{\text{noise}})$, has the entropy

$$H(S_{\text{noise}}) = H(S) + H(R_{\text{noise}}) = H(S) - mI(U, S), \tag{14}$$

where $H(S)$ is the entropy of the signal distribution. Equation 14 states that the magnitude of noise imposed by encoding increases with the variability of the input signal, which is caused by the limited capacity of the channel. Technically, the capacity of the channel can be elevated only by decreasing the internal noise.

Uniforming of datasets and the approximation of the transformation function

Uniforming of datasets is a direct approximation of Equations 12 and 13. Let $\Psi = \{\psi_i \in \mathbb{R}^2\}_{i=1,2,\ldots,N}$ be a two-dimensional dataset. Equation 12 leads directly to the uniforming of the first component by

$$f^1_\Psi(\psi^1_i) = \frac{1}{N}\sum_{j=1}^{N} H(\psi^1_i - \psi^1_j). \tag{15}$$

$f^1_\Psi(\psi^1_i)$ can be regarded as the approximation of f^1 at the sampling points ψ^1_i. The direct approximate of Equation 13 is

$$f^2_\Psi(\psi^1_i, \psi^2_i) = \frac{1}{\sum_{j=1}^{N} \delta_{\psi^1_i,\psi^1_j}} \sum_{j=1}^{N} \delta_{\psi^1_i,\psi^1_j} H(\psi^2_i - \psi^2_j).$$

However, even for dense datasets the number of data points exactly on the same line ($\delta_{\psi^1_i,\psi^1_j} = 1$) can be very small leading to an ill-approximated function f^2. Therefore, we introduce a weighting procedure to slightly smooth the data points around the line. Introducing the smoothing kernel $G(x, \sigma) = \exp(-x^2/2\sigma^2)$, we define

$$f^2_\Psi(\psi^1_i, \psi^2_i) = \frac{1}{\sum_{j=1}^{N} G(f^1_\Psi(\psi^1_i) - f^1_\Psi(\psi^1_j), \sigma)} \sum_{j=1}^{N} G(f^1_\Psi(\psi^1_i) - f^1_\Psi(\psi^1_j), \sigma) H(\psi^2_i - \psi^2_j). \tag{16}$$

The width σ of the smoothing kernel depends on the number N of the dataset. For our study ($N > 20\,000$) we set $\sigma = 0.0037$.

The approximation F_Ψ of the transformation function F can now be accomplished by linear interpolation using $f^1_\Psi(\psi^1_i)$ and $f^2_\Psi(\psi^1_i, \psi^2_i)$ as sample points.

Generating a new random sample from the transformation function F_Ψ

The approximated transformation function F_Ψ is a two-dimensional version of a cumulative function representing a PDF P_Ψ (see also Equation 11). The dataset Ψ is a random sample from that PDF. Since F_Ψ is invertible it is possible to generate in turn a new sample from the PDF P_Ψ. This can be simply done by generating a uniformly distributed random sample $U \in [0, 1] \times [0, 1]$. The new random sample $\tilde{\Psi}$ from P_Ψ can now be calculated by $\tilde{\Psi} = F^{-1}_\Psi(U)$.

Modeling the response of a neuron population, decoding scheme, and noise

The neuron population is constructed by uniformly covering the space $[0, 1] \times [0, 1]$ with N circular gaussian receptive fields with the following tuning curves:

$$r^i_m(s) = r_0 \exp(-(|2\pi(c^i_1 - s_1)|^2_{\text{circ}})/2(2\pi w)^2) \exp(-(c^i_2 - s_2))^2/2w^2),$$
$$s = (s_1, s_2) \in [0, 1] \times [0, 1], \tag{17}$$

(see also Section 'The construction of efficiently encoding neuron populations'), where $(c^i_1, c^i_2) \in [0, 1] \times [0, 1]$ marks the centers of the receptive fields, and w steers the width of the receptive field. For a neuron population with $N = n \times n$ neurons we arrange the receptive field centers by $c^i_1 = (\tau/n) + (1/2n)$; $c^i_2 = (\gamma/n) + (1/2n)$; $\tau, \gamma = 0, \ldots, n-1$, $i = n\gamma + \tau + 1$. Furthermore, we set $r_0 = 10$, and $w = 0.1$. We assume that the variability of the actual firing rate of each neuron obeys the homogenous Poisson model such that the probability $P(r^i|s)$ of the firing rate $r^i \in \mathbb{N}$ elicited by the stimulus s is given by

$$P(r^i|s) = \frac{(r^i_m(s))^{r^i}}{r^i!} \exp(-r^i_m(s)).$$

Since all neurons are assumed to fire independently the probability that a stimulus s elicits the response vector $r = (r_1, r_2, \ldots, r_n)$ is

$$P(r|s) = \prod_{i=1}^{N} P(r^i|s). \tag{18}$$

Now, the signal s can be retrieved from the response vector r with different decoding schemes. The maximum *a posteriori* (MAP) method finds an estimation s_{est} by maximizing $P(s|r)$ for a population response r elicited by the actual input $s_{\text{act}} \in [0, 1] \times [0, 1]$. Assuming $P(s) = 1$, $\forall s \in [0, 1] \times [0, 1]$, maximizing $P(s|r)$ is equivalent to maximizing $P(r|s)$.

We investigated the behavior of noise $R_{\text{noise}} = \{s_{\text{est}} - s_{\text{act}}\}$ and the entropy of noise $H(R_{\text{noise}})$ depending on the number N of neurons for all three estimation methods. To estimate $H(R_{\text{noise}})$ we generated for 40 randomly selected samples $s_{\sigma_{\text{act}}} \in [0, 1] \times [0, 1]$ a distribution $S_{\sigma_{\text{est}}} \subset [0, 1] \times [0, 1]$ of 1000 samples by subsequently eliciting population responses for the same stimulus $s_{\sigma_{\text{act}}}$ and retrieving the stimulus by the above method. From each estimated sample $S_{\sigma_{\text{est}}}$ the entropy $H(S_{\sigma_{\text{est}}})$ is estimated by the k-nearest neighbor-method (Kozachenko and Leonenko 1987) with $k = 10$. Although all $H(S_{\sigma_{\text{est}}})$ have approximately the same value, which means that R_{noise} does not depend on s_{act_σ}, the mean of all entropies $H(S_{\sigma_{\text{est}}})$ eventually gives the estimated noise entropy $H(R_{\text{noise}})$. The dependence of noise entropy $H(R_{\text{noise}})$ can be well described by the function

$$H(R_{\text{noise}})(N) = -2.05 - \sqrt{2}\,\log(N) \tag{19}$$

(see Figure 15). Furthermore, we proved that the noise $H(R_{\text{noise}})$ is approximately Gaussian by estimating the negentropies, which is the difference between the entropy of a Gaussian due to the standard deviation and the actual entropy of the concerning distribution, of the $S_{i_{\text{est}}}$. It turns out that these negentropies are vanishing for all samples s.

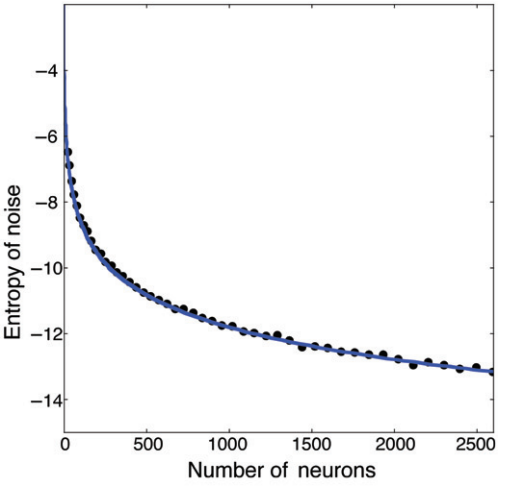

Figure 15. Noise entropy $H(R_{\text{noise}})$ for the maximum posterior estimator depending on the number of neurons, The variability of the population response is implemented by a homogeneous Poisson model. Black dots mark the estimated noise entropies for the respective numbers of neuron. The blue curve depicts the fit $H(R_{\text{noise}})(N) = -2.05 - \sqrt{2}\,\log(N)$.