Published in: Proceedings of the IEEE International Workshop on Biologically Motivated Computer Vision, Seoul, Korea, May 2000.
Lecture Notes in Computer Science 1811:492-501.

The Role of Natural Image Statistics in Biological Motion Estimation

Ron O. Dror[1], David C. O'Carroll[2], and Simon B. Laughlin[3]

[1] MIT, 77 Massachusetts Ave., Cambridge, MA 02139, USA, rdror@mit.edu
[2] University of Washington, Box 351800, Seattle, WA 98195-1800, USA
[3] University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

Abstract. While a great deal of experimental evidence supports the Reichardt correlator as a mechanism for biological motion detection, the correlator does not signal true image velocity. This study examines the accuracy with which physiological Reichardt correlators can provide velocity estimates in an organism's natural visual environment. Both simulations and analysis show that the predictable statistics of natural images imply a consistent correspondence between mean correlator response and velocity, allowing the otherwise ambiguous Reichardt correlator to act as a practical velocity estimator. A computer vision system may likewise be able to take advantage of natural image statistics to achieve superior performance in real-world settings.

1 Introduction

The Reichardt correlator model for biological motion detection [15] has gained widespread acceptance in the invertebrate vision community. This model, which is mathematically equivalent to the spatiotemporal energy models popular for vertebrate motion detection [2], has also been applied to explain motion detection in humans, birds, and cats [21, 22, 7]. After forty years of physiological investigation, however, a fundamental issue raised by Reichardt and his colleagues remains unanswered. While both insects and humans appear capable of estimating image velocity [18, 12], the output of a basic Reichardt correlator provides an inaccurate, ambiguous indication of image velocity. Correlator response to sinusoidal gratings depends on contrast (brightness) and spatial frequency (shape) as well as velocity; since the correlator is a nonlinear system, response to a broad-band image may vary erratically as a function of time. Some authors have concluded that velocity estimation requires either collections of differently tuned correlators [2], or an alternative motion detection system [18].

Before discarding the uniquely tuned Reichardt correlator as a velocity estimator, we consider the behavior of a physiologically realistic correlator in a natural environment. Previous experimental and modeling studies have typically focused on responses to laboratory stimuli such as sinusoidal or square gratings. We examine the responses of a Reichardt correlator to motion of natural broad-band images ranging from forests and animals to offices and city streets. In simulations, the correlator functions much better as a velocity estimator for

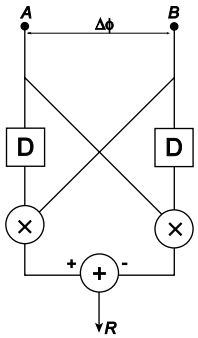

Fig. 1. A minimal Reichardt correlator.

motion of real-world imagery than for motion of traditional gratings. We develop a mathematical relationship between image power spectra and correlator response which shows that a system based on Reichardt correlators functions well in practice because natural images have predictable statistics and because the biological system is optimized to take advantage of these statistics. While this work applies generally to Reichardt correlators and mathematically equivalent models, we have chosen the fly as a model organism for computational simulations and experiments due to the abundance of behavioral, anatomical, and electrophysiological data available for its motion detection system.

The implication for machine vision is that the extensive recent body of work on statistics of natural images and image sequences (e.g., [8, 3, 17]) can be exploited in a computer vision system. Such a system may be superior to existing systems in practice despite inferior performance in simple test environments.

2 Correlator Response to Narrow-band Image Motion

Figure 1 shows a simplified version of the correlator model. Receptors A and B are separated by an angular distance $\Delta\phi$. The signal from A is temporally delayed by the low-pass filter D before multiplication with the signal from B. This multiplication produces a positive output in response to rightward image motion. In order to achieve similar sensitivity to leftward motion and in order to cancel excitation by stationary stimuli, a parallel delay-and-multiply operation takes place with a delay on the opposite arm. The outputs of the two multiplications are subtracted to give a single time-dependent correlator output R.

Although the correlator is nonlinear, its response to sinusoidal stimuli is of interest. If the input is a sinusoidal grating containing only a single frequency component, the oscillations of the two subunits cancel and the correlator pro-

duces a constant output.[1] If the delay filter D is first-order low-pass with time constant τ, as in most modeling studies (e.g., [6]), a sinusoid of amplitude C and spatial frequency f_s traveling to the right at velocity v produces an output

$$R(t) = \frac{C^2}{2\pi\tau} \frac{f_\mathrm{t}}{f_\mathrm{t}^2 + 1/(2\pi\tau)^2} \sin(2\pi f_\mathrm{s} \Delta\phi) \ , \tag{1}$$

where $f_\mathrm{t} = f_\mathrm{s} v$ is the temporal frequency of the input signal [6]. The output level depends separably on spatial and temporal frequency. At a given spatial frequency, the magnitude of correlator output increases with temporal frequency up to an optimum $f_{t,opt} = \frac{1}{2\pi\tau}$, and then decreases monotonically as velocity continues to increase. Output also varies with the square of C, which specifies grating brightness or, in the presence of preprocessing stages, grating contrast.

3 Correlator Response to Broad-band Images

Since the correlator is a nonlinear system, its response to a generic stimulus cannot be represented as a sum of responses to sinusoidal components of the input. In particular, the response to a broad-band image such as a natural scene may vary erratically with time.

3.1 Evaluation of Correlator Performance

In order to compare the performance of various velocity estimation systems, one must first establish a quantitative measure of accuracy. Rather than attempt to measure the performance of a motion detection system as a single number, we quantify two basic requirements for an accurate velocity estimation system:

1. Image motion at a specific velocity should always produce the same response.
2. The response to motion at a given velocity should be unambiguous; that is, it should differ from the response to motion at other velocities.

We restrict the range of potential input stimuli by focusing on responses to rigid, constant-velocity motion as observed by an eye undergoing rotational motion.

Given a large image moving at a particular constant velocity, consider an array of identically oriented correlators sampling the image at a dense grid of points in space and time. Define the mean response value $\overline{R}$ as the average of the ensemble outputs, and the relative error as the standard deviation of the ensemble divided by the mean response value. We call the graph of $\overline{R}$ as a function of velocity the velocity response curve. In order to satisfy requirement 1, different images should have similar velocity response curves and relative error should remain small. Requirement 2 implies that the velocity response curve should be monotonic in the relevant range of motion velocities.

[1] A physical luminance grating must have positive mean luminance, so it will contain a DC component as well as an oscillatory component. In this case, the output will oscillate about the level given by (1).

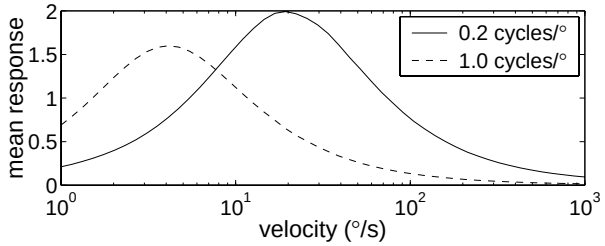

Fig. 2. Velocity response curves for the simple correlator model in response to sinusoidal gratings of two different spatial frequencies. Units on the vertical axis are arbitrary. The correlator in this simulation had a first-order low-pass delay filter with $\tau = 35$ ms, which matches the temporal frequency tuning observed experimentally in typical large flies such as *Calliphora*, *Eristalis*, and *Volucella* [9]. We set the inter-receptor angle to 1.08°, near the center of the physiologically realistic range for flies. These parameter choices do not critically influence our qualitative results.

Figure 2 shows velocity response curves for two simulated gratings of different spatial frequencies. The curves for the two gratings differ significantly, so that mean response level indicates velocity only if spatial frequency is known. In addition, the individual velocity response curves peak at low velocities, above which their output is ambiguous.

3.2 Simulation with Natural Images

One can perform similar simulations with natural images. In view of the fact that the characteristics of "natural" images depend on the organism in question and its behavior, we worked with two sets of images. The first set consisted of panoramic images photographed from favored hovering positions of the hoverfly *Episyrphus balteatus* in the woods near Cambridge, U.K. The second set of photographs, acquired by David Tolhurst, includes a much wider variety of imagery, ranging from landscapes and leaves to people, buildings, and an office [19]. Figure 3 displays images from both sets. We normalized each image by scaling the luminance values to a mean of 1.0, both in order to discount differences in units between data sets and to model photoreceptors, which adapt to the mean luminance level and signal the contrast of changes about that level [11].

Figure 4A shows velocity response curves for the images of Fig. 3. The most notable difference between the curves is their relative magnitude. When the curves are themselves normalized by scaling so that their peak values are equal (Fig. 4B), they share not only their bell shape, but also nearly identical optimal velocities. We repeated these simulations on most of the images in both sets, and found that while velocity response curves for different images differ significantly in absolute magnitude, their shapes and optimal velocities vary little.

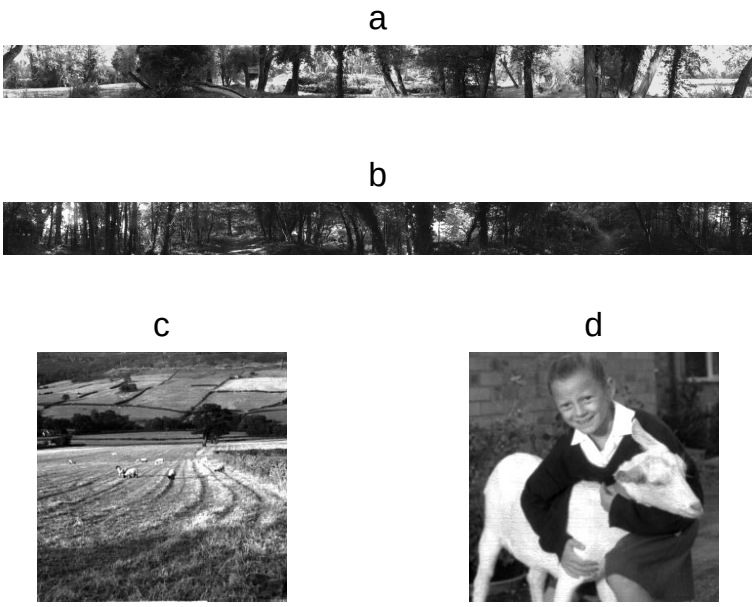

Fig. 3. Examples of the natural images used in simulations throughout this work. Images (a) and (b) are panoramic images from locations where *Episyrphus* chooses to hover. Images (c) and (d) are samples of the image set acquired by David Tolhurst. Details of image acquisition can be found in [5, 19].

This empirical similarity implies that if the motion detection system could normalize or adapt its response to remove the difference in magnitude between these curves, then the spatially or temporally averaged mean correlator response would provide useful information on image velocity relatively independent of the visual scene. One cannot infer velocity from the mean response of a correlator to a sinusoidal grating, on the other hand, if one does not know the spatial frequency in advance.

3.3 Mathematical Analysis of Mean Response to Broad-band Images

This section develops a general mathematical relationship between the power spectrum of an image and mean correlator response, explaining the empirical similarities in shape and differences in magnitude between velocity response curves for different images. Natural images differ from sinusoidal gratings in that they possess energy at multiple non-zero spatial frequencies, so that they produce broad-band correlator input signals. As an image moves horizontally across a horizontally-oriented correlator, one row of the image moves across the two correlator inputs. One might think of this row as a sum of sinusoids representing its Fourier components. Because of the nonlinearity of the multiplication

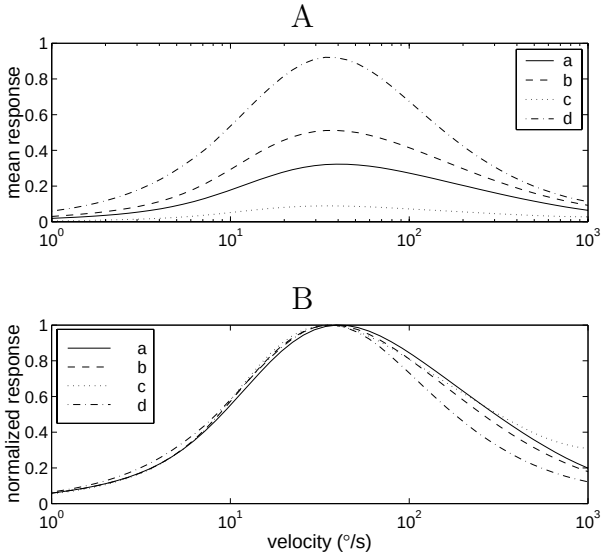

Fig. 4. Response of the simple correlator model to the natural images shown in Fig. 3. (A) Velocity response curves showing mean response to motion at different velocities, computed as in Fig. 2. (B) The same curves, normalized so that their maximum values are identical. Peak response velocities range from 35–40°/s.

operation, the correlator output in response to the moving image will differ from the sum of the responses to the individual sinusoidal components. In particular, the response to a sum of two sinusoids of different frequencies f_1 and f_2 consists of the sum of the constant responses predicted by (1) to each sinusoid individually, plus oscillatory components of frequencies $f_1 + f_2$ and $|f_1 - f_2|$. Sufficient spatial or temporal averaging of the correlator output will eliminate these oscillatory components. The correlator therefore exhibits pseudolinearity or linearity in the mean [14], in that the mean response to a broad-band image is equal to the sum of the responses to each sinusoidal input component.

This pseudolinearity property implies that the mean response of a simple Reichardt correlator to a single row of an image depends only on the power spectrum of that row. Using (1) for correlator response to a sinusoid and the fact that $f_\mathrm{t} = f_\mathrm{s} v$, we can write the mean correlator output as

$$\overline{R} = \frac{1}{2\pi\tau} \int_0^\infty P(f_\mathrm{s}) \frac{f_\mathrm{s} v}{(f_\mathrm{s} v)^2 + 1/(2\pi\tau)^2} \sin(2\pi f_\mathrm{s} \Delta\phi) df_\mathrm{s} \ , \tag{2}$$

where $P(f_\mathrm{s})$ represents the power spectral density of one row of the image at spatial frequency f_s. Each velocity response curve shown in Fig. 4 is an average of the mean outputs of correlators exposed to different horizontal image rows with potentially different power spectra. This average is equivalent to the response of

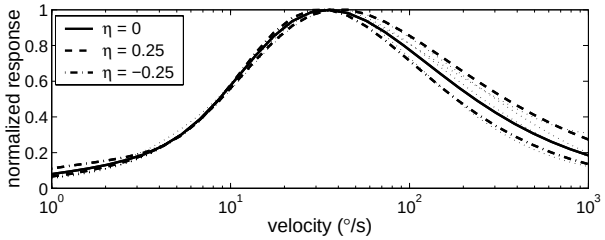

Fig. 5. Velocity response curves computed theoretically using (2), assuming row power spectra $P(f_s)$ of the form $f_s^{-(1+\eta)}$ for several values of η. Model parameters are as in Fig. 2. Simulated velocity response curves from Fig. 4B are shown in thin dotted lines for comparison. All curves have been normalized to a maximum value of 1.0. The predicted peak response velocities are 32, 35, and 40°/s for $\eta = -0.25$, 0, and 0.25, respectively.

the correlator to a single row whose power spectrum $P(f_s)$ is the mean of the power spectra for all rows of the image.

If $P(f_s)$ were completely arbitrary, (2) would provide little information about the expected shape of the velocity response curve. A large body of research suggests, however, that power spectra of natural images are highly predictable. According to a number of studies involving a wide range of images, the two-dimensional power spectra are generally proportional to $f^{-(2+\eta)}$, where f is the modulus of the two-dimensional spatial frequency and η is a small constant (e.g., [8, 19]). If an image has an isotropic two-dimensional power spectrum proportional to $f^{-(2+\eta)}$, the one-dimensional power spectrum of any straight-line section through the image is proportional to $f^{-(1+\eta)}$.

Overall contrast, which determines overall amplitude of the power spectrum, varies significantly between natural images and between orientations [20]. The best value of η also depends on image and orientation, particularly for images from different natural environments. Van der Schaaf and van Hateren [20] found, however, that a model which fixes $\eta = 0$ while allowing contrast to vary suffers little in its fit to the data compared to a model which allows variation in η.

The similarities in natural image power spectra lead to predictable peak response velocities and to similarities in the shapes of the velocity response curves for different images. Figure 5 shows velocity response curves predicted from hypothetical row power spectra $P(f_s) = f_s^{-1}$, $f_s^{-1.25}$, and $f_s^{-0.75}$, corresponding to $\eta = 0$, 0.25, and -0.25, respectively. The theoretical curves match each other and the simulated curves closely below the peak response value; in this velocity range, the velocity response is insensitive to the value of the exponent in the power spectrum.

Contrast differences between images explain the primary difference between the curves, their overall amplitude. Figure 6 shows horizontal power spectral densities for the images of Fig. 3, computed by averaging the power spectral densities of the rows comprising each image. On log-log axes, the spectra ap-

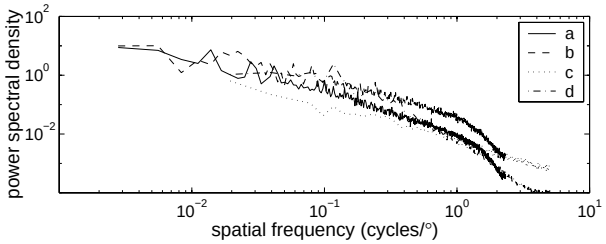

Fig. 6. Horizontal power spectral densities of the images in Fig. 3. Each spectrum is an average of power spectral densities of the rows comprising the image. Images (a) and (b) roll off in power at frequencies above 1.2 cycles/° due to averaging in the image acquisition process, but optical blur effects in the fly's eye reject almost all spatial frequency content above 1 cycle/°.

proximate straight lines with slopes close to -1, although the spectrum of image (d) has noticeable curvature. The relative magnitudes of the spectra correspond closely to the relative magnitudes of the velocity response curves of Fig. 4, as predicted by (2). Differences in the magnitude of the velocity response curves correspond to differences in overall contrast, except that image (d) has the largest response even though its contrast is only larger than that of (b) for frequencies near 0.1 cycles/°. This reflects the fact that some spatial frequencies contribute more than others to the correlator response.

In order to use mean correlator response as a reliable indicator of velocity, the visual system needs to compensate for these contrast variations. One possibility is that contrast saturation early in the motion detection pathway eliminates significant differences in contrast. Alternatively, some form of contrast normalization akin to that observed in vertebrate vision systems [10] may work to remove contrast differences between images. Ideal localized contrast normalization would remove the dependence of correlator response on the spatial frequency of a sinusoidal grating [1], but this dependence has been documented experimentally, and our work suggests that a more global form of contrast normalization is likely. Our experimental results [13] confirm the relationships between the power spectrum and the velocity response curve predicted in this section and suggest that the response of wide-field neurons reflects image velocity consistently even as image contrast changes.

3.4 Limitations and Further Work

While the simple correlator model of Fig. 1 produces more meaningful estimates of velocity for natural images than for arbitrary sinusoids, it suffers from two major shortcomings. First, the standard deviation of the correlator output is huge relative to its mean, with relative error values ranging from 3.3 to 76 for the images and velocity ranges of Fig. 4. Second, mean correlator response for most natural images peaks at a velocity of 35 to 40°/s. While this velocity sig-

nificantly exceeds the peak response velocity of 19.6°/s for a sinusoidal grating of optimal spatial frequency, it still leads to potential ambiguity since flies may turn and track targets at velocities up to hundreds of degrees per second. In further work [5, 4], we found that a more physiologically realistic correlator model raises velocity response and lowers relative error dramatically through the inclusion of experimentally described mechanisms such as input prefiltering, output integration, compressive nonlinearities, and adaptive effects. Equation 2 generalizes naturally to predict the quantitative and qualitative effects of spatial and temporal prefiltering. This work examines only responses to constant velocity rigid motion; further work should consider natural image sequences.

4 Conclusions

While natural images appear more complicated than the gratings typically used in laboratory experiments and simulations, Reichardt correlators respond more predictably to motion of natural images than to gratings. The general structure and detailed characteristics of the physiological correlator suggest that it evolved to take advantage of natural image statistics for velocity estimation. While we worked with models based on data from insect vision, these conclusions also apply to models of vertebrate vision such as the Elaborated Reichardt Detector [21] and the spatiotemporal energy model [2], both of which are formally equivalent to the Reichardt correlators discussed here.

These results could be applied directly to hardware implementations of correlator-based motion-detection systems (e.g., [16]). A more important implication is that a machine vision system designed to perform a task involving real-world imagery would do well to take advantage of recent results in the field of natural image statistics. These results include statistics of image sequences [3] and extend well beyond power spectra [17]. A Bayesian approach to computer vision should take these statistics into account, as does the biological motion detection system.

Acknowledgments

We would like to thank David Tolhurst for sharing his set of images and Miranda Aiken for recording the video frames which formed the panoramic images. Rob Harris, Brian Burton, and Eric Hornstein contributed valuable comments. This work was funded by a Churchill Scholarship to ROD and by grants from the BBSRC and the Gatsby Foundation.

References

1. E. H. Adelson and J. R. Bergen. The extraction of spatio-temporal energy in human and machine vision. In *Proceedings from the Workshop on Motion: Representation and Analysis*, pages 151–55, Charleston, SC, 1986.

2. E. H. Adelson and J.R. Bergen. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A*, 2:284–99, 1985.
3. D. W. Dong and J. J. Atick. Statistics of natural time-varying images. *Network: Computation in Neural Systems*, 6:345–58, 1995.
4. R. O. Dror. Accuracy of velocity estimation by Reichardt correlators. Master's thesis, University of Cambridge, Cambridge, U.K., 1998.
5. R. O. Dror, D. C. O'Carroll, and S. B. Laughlin. Accuracy of velocity estimation by Reichardt correlators. Submitted.
6. M. Egelhaaf, A. Borst, and W. Reichardt. Computational structure of a biological motion-detection system as revealed by local detector analysis in the fly's nervous system. *J. Opt. Soc. Am. A*, 6:1070–87, 1989.
7. R. C. Emerson, M. C. Citron, W. J. Vaughn, and S. A. Klein. Nonlinear directionally selective subunits in complex cells of cat striate cortex. *J. Neurophysiology*, 58:33–65, 1987.
8. D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4:2379–94, 1987.
9. R. A. Harris, D. C. O'Carroll, and S. B. Laughlin. Adaptation and the temporal delay filter of fly motion detectors. *Vision Research*, 39:2603–13, 1999.
10. D. J. Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9:181–97, 1992.
11. S. B. Laughlin. Matching coding, circuits, cells and molecules to signals: general principles of retinal design in the fly's eye. *Prog. Ret. Eye Res.*, 13:165–95, 1994.
12. Suzanne P. McKee, Gerald H. Silverman, and Ken Nakayama. Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision Research*, 26:609–19, 1986.
13. D. C. O'Carroll and R. O. Dror. Velocity tuning of hoverfly HS cells in response to broad-band images. In preparation.
14. T. Poggio and W. Reichardt. Visual control of orientation behaviour in the fly. Part II. Towards the underlying neural interactions. *Quarterly Reviews of Biophysics*, 9:377–438, 1976.
15. W. Reichardt. Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In A. Rosenblith, editor, *Sensory Communication*, pages 303–17. MIT Press and John Wiley and Sons, New York, 1961.
16. R. Sarpeshkar, W. Bair, and C. Koch. An analog VLSI chip for local velocity estimation based on Reichardt's motion algorithm. In S. Hanson, J. Cowan, and L. Giles, editors, *Advances in Neural Information Processing Systems*, volume 5, pages 781–88. Morgan Kauffman, San Mateo, 1993.
17. E.P. Simoncelli. Modeling the joint statistics of images in the wavelet domain. In *Proc SPIE, 44th Annual Meeting*, volume 3813, Denver, July 1999.
18. M. V. Srinivasan, S. W. Zhang, M. Lehrer, and T. S. Collett. Honeybee navigation *en route* to the goal: visual flight control and odometry. *J. Exp. Biol.*, 199:237–44, 1996.
19. D. J. Tolhurst, Y. Tadmor, and T. Chao. Amplitude spectra of natural images. *Ophthalmology and Physiological Optics*, 12:229–32, 1992.
20. A. van der Schaaf and J. H. van Hateren. Modelling the power spectra of natural images: statistics and information. *Vision Research*, 36:2759–70, 1996.
21. J. P. H. van Santen and G. Sperling. Elaborated Reichardt detectors. *J. Opt. Soc. Am. A*, 2:300–21, 1985.
22. F. Wolf-Oberhollenzer and K. Kirschfeld. Motion sensitivity in the nucleus of the basal optic root of the pigeon. *J. Neurophysiology*, 71:1559–73, 1994.