A Computational Model of Biological Motion Processing.

This section contains a general description of the model employed in the study. A full description is given by Johnston et al (1999). We initially describe a version that operates in only two dimensions (i.e. one spatial dimension, *x*, and one temporal dimension, *t*).

The model is derived from the gradient approach in which velocity is calculated by dividing the temporal derivative of image brightness by the spatial derivative of image brightness. Because partial differentiation of an image followed by filtering is equivalent to applying a partially differentiated filter to the image (Bracewell, 1965), a gradient model may be realised by applying pairs of filters to an image, one of which is the temporal derivative and one the spatial derivative of the same filter kernel. Local velocity can then be calculated by taking the ratio of the outputs of the two filters.

One potential problem is that, when the output of the spatial filter is zero, velocity is undefined. In order to condition the velocity ratio calculation, one can include additional measures. Rather than thinking of a single filter kernel, one can think of a vector of filter kernels. By taking the spatial derivatives of each of the filter kernels we derive one vector of filters, and by taking the temporal derivatives we derive another. A least squares estimate of velocity can then be calculated as the ratio of dot products:

$$velocity = \frac{X \cdot T}{X \cdot X} \qquad \text{Equation 1}$$

Where *X* is the vector of filter outputs from the spatially differentiated filters, and *T* is a vector of filter outputs from the temporally differentiated filters. In this case, for the denominator to equal zero, all of the measures in *X* must equal zero. If we increase the number of measures then the probability of this occurring is reduced.

Here, however, one is faced with the problem of how to construct the vector of filter kernels prior to differentiation. What filters should be used, and how should they be weighted relative to one another? One way to solve this problem is to utilise a proposal put forward by Koenderink & van Doorn (1987). In this framework, local image structure is represented as a truncated Taylor series expansion. In other words, image structure in a local region is represented by the weighted outputs of a series of filters applied at a point in the image. These filters are generated from a single filter (the *blur kernel*) by progressively increasing the order of the spatial and temporal differential operators applied to the kernel.

From the Taylor series expansion, one can estimate image brightness at a distance from the point at which the measures are taken. The weights that are attached to the various derivatives are dependent upon the direction and length of the vector joining the point at which the measures are taken and the point at which we wish to estimate image brightness. Therefore, for any point in the image we can construct a vector of weighted filter functions which will serve as the initial vector of filter kernels. The calculation of velocity uses this property of the Taylor series to integrate over a region surrounding the point at which the bank of filters is applied.

The 2D space-time version of the model summarised so far, has been described by Johnston et al (1992) and Johnston & Clifford (1995). In the current study we use a version that has been extended to accommodate 3 dimensions (i.e. two spatial dimensions, *x* and *y*, and one temporal dimension, *t*). To generate our vector of filter kernels from a single blur kernel, we differentiate the kernel in increasing orders of *x*, *y*, and *t*. As in the 2D version of the model, the blur kernel is a Gaussian in space and log time.

In the 3D version of the model, the reference frame (i.e. the *x* and *y* axes defining the local direction of spatial differentiation) is rotated through a number of orientations with respect to the input image. A total of 24 orientations (evenly spaced over 360 degrees) are employed. For each orientation, three vectors of filters are created by differentiating the vector of filter kernels with respect to *x*, *y* and *t*. Coupled with the rotation of the reference frame, this produces a population of filters that are tuned to different orientations and spatial frequencies and may show either transient or sustained temporal properties. From the measures derived by applying the three vectors of filter functions to the image, we calculate the following 4 speed related measures: speed, orthogonal speed, inverse speed and orthogonal inverse speed:

$$\textit{speed} = \frac{X \cdot T}{X \cdot X}\cos^2\phi = \frac{X \cdot T}{X \cdot X}\left(1+\left(\frac{X \cdot Y}{X \cdot X}\right)^2\right)^{-1} \qquad \text{Equation 2}$$

$$\textit{orthogonal speed} = \frac{Y \cdot T}{Y \cdot Y}\sin^2\phi = \frac{Y \cdot T}{Y \cdot Y}\left(1+\left(\frac{X \cdot Y}{Y \cdot Y}\right)^2\right)^{-1} \qquad \text{Equation 3}$$

$$\textit{inverse speed} = \frac{X \cdot T}{T \cdot T} \qquad \text{Equation 4}$$

$$\textit{orthogonal inverse speed} = \frac{Y \cdot T}{T \cdot T}\ , \qquad \text{Equation 5}$$

where *X* is the vector of filter outputs from the *x* differentiated filters, *Y* is the vector of filter outputs from the *y* differentiated filters and *T* is the vector of filter outputs from the temporally differentiated filters. Speed ($X \cdot T / X \cdot X$) and orthogonal speed ($Y \cdot T / Y \cdot Y$) measures are ill conditioned when there is no variation over *x* and *y* respectively. To prevent this from degrading the final velocity estimate, speed and orthogonal speed are both conditioned by measures of the angle of image structure relative to the reference frame (Φ). From Equation 2, it can be seen that as $X \cdot X$ decreases, $(X \cdot Y / X \cdot X)^2$ increases more rapidly than the speed term, thereby forcing the product to zero as ($X \cdot T / X \cdot X$) approaches infinity. The inverse speed measures however become infinite when the temporal derivative is close to zero, which is appropriate.

As the four speed related measures are calculated from different combinations of different filter outputs, they may all thought of as different estimates that are related to local speed. The inclusion of additional image measures to increase robustness is one of the core design criteria that runs through the model.

By taking each of the speed related measures indexed by the orientations through which the reference frame is rotated, we can form the vectors, $\widehat{s}_{\|}$ (speed), $\widehat{s}_{\perp}$ (orthogonal speed), $\breve{s}_{\|}$ (inverse speed) and $\breve{s}_{\perp}$ (orthogonal inverse speed). The final speed estimate is calculated as the square root of the ratio of determinants:

$$speed^2 = \frac{\begin{vmatrix} \widehat{s}_{\|} \cdot F_{\|} & \widehat{s}_{\|} \cdot F_{\perp} \\ \widehat{s}_{\perp} \cdot F_{\|} & \widehat{s}_{\perp} \cdot F_{\perp} \end{vmatrix}}{\begin{vmatrix} \widehat{s}_{\|} \cdot \breve{s}_{\|} & \widehat{s}_{\|} \cdot \breve{s}_{\perp} \\ \widehat{s}_{\perp} \cdot \breve{s}_{\|} & \widehat{s}_{\perp} \cdot \breve{s}_{\perp} \end{vmatrix}} \qquad \text{Equation 6}$$

where $F_{\|}$ and $F_{\perp}$ are vectors containing the cosines and sines of the angles through which the reference frame has been rotated.

In the case of the motion of a simple pattern such as, for example, a translating sine wave grating, the denominator of Equation 6 takes a value of one. Measures of speed and orthogonal speed vary sinusoidally with angle of reference frame, with orthogonal speed lagging behind speed by a quarter of a cycle. The numerator shown in Equation 6 works by taking a measure of the amplitude of the distribution of speed measures that is combined across both speed and orthogonal speed. The denominator (and its measures of inverse speed) are included to stabilise the final speed estimation. This can be shown to be particularly important in the case where motion occurs in the presence of static pattern (see Johnston et al (1999) for a full analysis).

Direction of motion is extracted by calculating a measure of phase that is combined across all 4 speed related measures:

$$direction = \tan^{-1}\left(\frac{(\breve{s}_{\|} + \widehat{s}_{\|}) \cdot F_{\perp} + (\breve{s}_{\perp} + \widehat{s}_{\perp}) \cdot F_{\|}}{(\breve{s}_{\|} + \widehat{s}_{\|}) \cdot F_{\|} - (\breve{s}_{\perp} + \widehat{s}_{\perp}) \cdot F_{\perp}}\right). \qquad \text{Equation 7}$$

When speed is large (and inverse speed is small) then direction is dominated by the speed measures, however when speed is small (and inverse speed is large) then the measure is dominated by inverse speed. The use of complementary and antagonistic speed and inverse speed measures should prove valuable in any system where small signals are likely to be affected by noise.

The model described above incorporates a number of distinct processing stages and integrates the outputs of oriented filters that have a range of spatial and temporal frequency tuning characteristics, consistent with the properties of cortical neurons. In summary, the model can be thought of as a stable and robust extension of the gradient algorithm. The model seeks to resolve problems of mathematical ill-conditioning by making multiple measures on the stimulus rather than by introducing arbitrary constants or thresholds. It should be noted that the model contains no non-linear preprocessing stages that seek to extract texture or features prior to motion analysis; the computation of speed and direction are based directly upon the output of filters applied to the input image.

Note: The passage shown above has been adapted from Benton, Johnston & McOwan (in press).

References.

Benton, C.P., Johnston, A. & McOwan, P.W. (In press). Computational modelling of interleaved first- and second-order motion sequences and translating 3f + 4f beat patterns. *Vision Research*.

Bracewell, R.N. (1965). The Fourier transform and its applications. *New York: McGraw-Hill*.

Johnston, A. & Clifford, C.W.G. (1995). Perceived motion of contrast modulated gratings: predictions of the multi-channel gradient model and the role of full-wave rectification. Vision Research, 35, 1109-1123.

Johnston, A., McOwan, P.W. & Buxton, H. (1992). A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proceedings of the Royal Society of London, B, 250*, 297-306

Johnston, A., McOwan, P.W. & Benton, C.P. (1999). Robust velocity computation from a biologically motivated model of motion perception. *Proceedings of the Royal Society of London, B, 266*, 509-518

Koenderink, J.J. & van Doorn, A.J. (1987). Representation of local-geometry in the visual system. *Biological Cybernetics, 55*, 367-375.