Supplementary material for May, K.A. & Zhaoping, L. "Li and Atick's theory of efficient binocular coding: A tutorial and mini-review"

APPENDIX A: Quantifying information

The amount of information that we gain from receiving a signal can be equated to how surprising the signal is. If we were already very sure what the signal would be, and the signal confirms our expectations, then we haven't learnt very much; but if the signal that we receive was unlikely, then we are surprised, and have learnt more. Surprise, or information, is therefore inversely related to the probability of the signal, $P(S)$. In information theory, the particular function we use is given by

$$\text{Surprise} = -\log_2 P(S). \tag{A1}$$

Thus, when $P(S) = 1$, Surprise $= 0$ (i.e. we knew what the signal would be, so have gained no information by receiving it), and Surprise approaches ∞ as $P(S)$ approaches 0 (i.e. the signal was very unlikely, so we have gained a lot of information by receiving it).

Shannon (1948) defined entropy, H, of a signal as the "expected" amount of surprise, i.e. the average surprise over repeated presentations of the signal. This is given by the sum of all the different possible signals, weighted by the probability of each signal:

$$H(S) = -\sum_S P(S)\log_2 P(S). \tag{A2}$$

Alternatively, for continuous signals, the entropy is given by

$$H(S) = -\int_S p(S)\log_2 p(S)dS, \tag{A3}$$

where $p(S)$ is the probability density of S. For mathematical convenience, we will usually take the signal to be continuous.

Entropy is measured in bits (binary digits). This is because, if the signal, S, contains n binary digits, then there are 2^n possible states, so if each state is equally likely, the probability of each state is $1/2^n$, and the entropy according to Equation (A2) is n, i.e. the number of bits in the signal.

Entropy can be considered to be the amount of information about the signal that we lack before receiving it. If the signal, S, is an external sensory signal, such as a pattern of light, the brain does not receive the signal directly; instead, it receives the output, O, of a neuron or set of neurons. Before receiving the neuronal output, O, the amount of information about S that we lacked was $H(S)$. After receiving O, the probability of the signal, S, is the conditional probability, $p(S \mid O)$; the amount of information about S that we still expect to lack is then given by the conditional entropy, which is defined analogously to $H(S)$:

$$H(S \mid O) = -\iint_{O\,S} p(O,S)\log_2 p(S \mid O)dSdO. \tag{A4}$$

The amount of information about S that we expect to gain by receiving O is given by the difference between what we lacked before receiving the signal and what we expect to lack after receiving the signal:

$$I(O;S) = H(S) - H(S \mid O). \tag{A5}$$

$I(O;S)$ is called the *mutual information*: it tells us how much information O provides about S.

APPENDIX B: Understanding the effect of ($\phi_- - \phi_+$) and sensitivity ratio g_+/g_- on binocularity

For a given g_+/g_-, binocularity is maximised when $K_+(x)$ and $K_-(x)$ are in quadrature phase, and minimised when their phase difference is 0° or 180°

When $K_+(x)$ and $K_-(x)$ are in quadrature phase, $|\phi_- - \phi_+| = 90°$, so $\cos(\phi_- - \phi_+) = 0$. Thus, from Equations (79) and (80), $O_R = O_L$, and $OBI = 1$: The neuron is fully binocular, regardless of the value of g_+/g_-. Equations (79) and (80) show that the source of the difference between O_R and O_L is the $2(g_+/g_-)\cos(\phi_- - \phi_+)$ term, which is positive in one equation and negative in the other. When $K_+(x)$ and $K_-(x)$ differ in phase by 0° or 180°, $\cos(\phi_- - \phi_+) = \pm 1$, so the magnitude of $2(g_+/g_-)\cos(\phi_- - \phi_+)$ is maximised, which maximises the difference between O_R and O_L. Thus, the OBI is minimised.

We can also understand these phenomena visually, by considering the vectors $\mathbf{v}_+$, $\mathbf{v}_-$, $\mathbf{v}_L$ and $\mathbf{v}_R$, which represent the sensitivities and phases of the corresponding receptive fields. Figures 3–5 show that, regardless of the value of g_+/g_-, when $\mathbf{v}_+$ and $\mathbf{v}_-$ are orthogonal (indicating that $K_+(x)$ and $K_-(x)$ are in quadrature phase), their sum and difference ($\mathbf{v}_R$ and $\mathbf{v}_L$) fall on the diagonals of rectangles with identical width and height, and are thus equal in length. Thus, $K_R(x)$ and $K_L(x)$ are equal in sensitivity, and the neuron is fully binocular. As $\mathbf{v}_+$ and $\mathbf{v}_-$ move away from orthogonality, they begin to point in similar or opposite directions. Thus their sum will increase or decrease in magnitude, while their difference does the opposite: One of $\mathbf{v}_R$ and $\mathbf{v}_L$ will increase in length while the other decreases. This means that $K_R(x)$ and $K_L(x)$ will start to differ in sensitivity, and the neuron will not be fully binocular. The difference between the lengths of $\mathbf{v}_R$ and $\mathbf{v}_L$ is maximised when $\mathbf{v}_+$ and $\mathbf{v}_-$ point in exactly the same or exactly opposite directions (i.e. $K_+(x)$ and $K_-(x)$ differ in phase by 0° or 180°): at this point, binocularity is minimised.

For a given phase difference, $\phi_- - \phi_+$, binocularity is minimised when $g_+/g_- = 1$

This is easiest to understand by considering the vectors $\mathbf{v}_+$, $\mathbf{v}_-$, $\mathbf{v}_L$ and $\mathbf{v}_R$. As noted above, when $\mathbf{v}_+$ and $\mathbf{v}_-$ are not orthogonal, they will tend to cancel each other out in either their sum ($\mathbf{v}_R$) or their difference ($\mathbf{v}_L$). The more similar the lengths of $\mathbf{v}_+$ and $\mathbf{v}_-$, the more complete this cancellation can be. The maximum cancellation occurs when they are the same length, i.e. $g_+/g_- = 1$ (Figure 3). In the extreme, when $K_+(x)$ and $K_-(x)$ differ in

phase by 0° or 180°, $\mathbf{v}_+$ and $\mathbf{v}_-$ point in either the same direction or opposite directions, so if $g_+/g_- = 1$, then either $\mathbf{v}_R$ or $\mathbf{v}_L$ has zero magnitude, and the neuron is completely monocular. For $g_+/g_- \neq 1$, cancellation can never be complete (e.g. Figures 4 and 5), so the neuron would be more binocular.

Alternatively, we can gain insights into the shape of the curves in Figure 8 by considering the equations. It is easiest to understand the case of $\phi_- - \phi_+ = 0$. In this case, Equations (79) and (80) simplify to

$$O_R = |g_+/g_- + 1| \tag{B1}$$

$$O_L = |g_+/g_- - 1|. \tag{B2}$$

When $g_+/g_- = 1$, $O_L = 0$, so the neuron is completely monocular. A similar argument applies in the case of $\phi_- - \phi_+ = 180°$.

To understand why the curves are symmetrical about the line $g_+/g_- = 1$ on the log axis in Figure 8, let us take O_R and O_L to be functions of $q = g_+/g_-$:

$$O_R(q) = \sqrt{q^2 + 1 + 2q\cos(\phi_- - \phi_+)} \tag{B3}$$

$$O_L(q) = \sqrt{q^2 + 1 - 2q\cos(\phi_- - \phi_+)}. \tag{B4}$$

Then we have

$$O_R(1/q) = \sqrt{(1/q)^2 + 1 + (2/q)\cos(\phi_- - \phi_+)} = (1/q)O_R(q) \tag{B5}$$

$$O_L(1/q) = \sqrt{(1/q)^2 + 1 - (2/q)\cos(\phi_- - \phi_+)} = (1/q)O_L(q). \tag{B6}$$

The *OBI* is unaffected if the same multiplier is applied to both O_R and O_L, as it will cancel out in Equation (76), and so the *OBI* is identical for $g_+/g_- = q$ and $g_+/g_- = 1/q$. Thus the *OBI* is identical for $\log(g_+/g_-) = \log(1) \pm \log(q)$, so on log axes, the curves are symmetrical about the line $g_+/g_- = 1$.

Reference

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal, 27*, 379-423, 623-656.