Stereoscopic depth adaptation from binocularly correlated versus anticorrelated noise: test of an efficient coding theory of stereopsis

Frederick A. A. Kingdom[1], Karl-Christopher Yared[1], Paul Hibbard[2], Keith May[2]

1. McGill Vision Research, Department of Ophthalmology, Montréal General Hospital, 1650 Cedar Ave., Rm. L11.112, Montréal, PQ, H3G 1A4, Canada

2. Department of Psychology, University of Essex, Wivenhoe Park, Essex CO4 3SQ, UK

Corresponding author: fred.kingdom@mcgill.ca

Keywords: Stereopsis, efficient coding, adaptation, binocular differencing channel

Abstract

Stereoscopic, or "3D" vision in humans is mediated by neurons sensitive to the disparities in the positions of objects in the two eyes' views. A disparity-sensitive neuron is typically characterized by its responses to left- and right-eye monocular signals, $\mathbf{S_L}$ and $\mathbf{S_R}$, respectively. However, it can alternatively be characterized by sensitivity to the sum of the two eyes' inputs, $\mathbf{S_+} = \mathbf{S_L} + \mathbf{S_R}$, and the difference, $\mathbf{S_-} = \mathbf{S_L} - \mathbf{S_R}$. Li and Atick's theory of efficient binocular encoding proposes that the $\mathbf{S_+}$ and $\mathbf{S_-}$ signals can be separately weighted to maximize the efficiency with which binocular information is encoded. This adaptation changes each neuron's sensitivity and preferred binocular disparity, resulting in predicted effects on the perceived stereoscopic depth of objects. To test these predictions, we measured the apparent depth of a random-dot stereogram with an 'in-front' target following adaptation to binocularly correlated or anti-correlated horizontally-oriented

grating stimuli, which reduce sensitivity to the $\mathbf{S}_+$ and $\mathbf{S}_-$ signals, respectively, but which contain no conventional stereo-depth signals. The anti-correlated noise adaptation made the target appear relatively closer to the background than the correlated noise adaptation, with differences of up to 60%. We show how this finding can be accommodated by a standard model of binocular disparity processing, modified to incorporate the binocular adaptation suggested by Li and Atick's theory.

Keywords: stereopsis, adaptation, efficient coding, binocular summing channel, binocular differencing channel

Introduction

Humans and animals are said to have binocular vision if they possess two spatially separated eyes that together provide a coherent image of the external world. Two eyes offer a range of advantages over one: for example a wider field of view, and, importantly, stereoscopic or "3D" vision. To achieve stereopsis, the visual system detects disparities in the positions of objects in the two eyes. Conventionally, stereo-sensitive neurons are described in terms of the left-eye and right-eye receptive fields, $\mathbf{K}_\mathbf{L}$ and $\mathbf{K}_\mathbf{R}$, which map out the neuron's sensitivity to each eye's image as a function of space. The neuron's binocular disparity tuning is determined by the offset between the left and right eye receptive fields in position or phase (Ohzawa, DeAngelis & Freeman, 1990; Anzai, Ohzawa & Freeman, 1999). However, a theory advanced by Li & Atick (1994) describes the neurons in terms of the summation and difference receptive fields, $\mathbf{K}_+$ and $\mathbf{K}_-$, which map out the neuron's sensitivity to the binocular sum and difference images, $\mathbf{S}_+$ and $\mathbf{S}_-$, respectively, as functions of space. Unlike the left and right eye images, $\mathbf{S}_+$ and $\mathbf{S}_-$ are abstract entities that the visual system is never directly exposed to ($\mathbf{S}_+$ is the sum of the two eyes' images,

and $\mathbf{S}_-$ is the difference between them); nevertheless, it can be useful to express the neuron's sensitivity to $\mathbf{S}_+$ and $\mathbf{S}_-$ as a function of space, just as we express the same neuron's sensitivity to the left and right eye images.

It is important to emphasize that, by describing the neuron in terms of its sensitivity to the $\mathbf{S}_+$ and $\mathbf{S}_-$ images, we are not proposing a fundamentally different neuronal model: we are taking the standard model of a neuron with left and right eye receptive fields, but describing it differently. The two descriptions (one in terms of $\mathbf{K}_L$ and $\mathbf{K}_R$, and the other in terms of $\mathbf{K}_+$ and $\mathbf{K}_-$) are exactly equivalent, and can be derived from each other (analogously, we can describe a linear neuron in terms of its sensitivity to different spots of light – the receptive field – or in terms of its sensitivity to different spatial frequencies – the spatial frequency tuning curve; these are not different models – they are different descriptions of the same model, which can be derived from each other). The reason that Li and Atick chose to describe the neurons in terms of their $\mathbf{K}_+$ and $\mathbf{K}_-$ receptive fields is that they propose that these receptive fields implement, respectively, binocular summation and differencing channels that are multiplexed on the same neuron. It is proposed that the sensitivity of each channel can be independently adjusted to maximize coding efficiency. As we shall see, this adjustment of channel sensitivity has important consequences for the neurons' left and right eye receptive fields, $\mathbf{K}_L$ and $\mathbf{K}_R$.

Mounting evidence supports the existence of both binocular summation and differencing channels (DeSilva & Bartley, 1930; Cohn & Lasley, 1976; Cohn, Leong & Lasley, 1981; Li & Atick, 1994; May, Zhaoping & Hibbard, 2012; Kingdom, 2011; Said & Heeger, 2013; Zhaoping, 2014; Yoonessi & Kingdom, 2009; May & Zhaoping, 2016, 2019; Henriksen & Read, 2016; Jennings & Kingdom, 2016; Georgeson et al., 2016), but their involvement in stereopsis has up to now been purely speculative (Li & Atick, 1994; Kingdom, 2011; Zhaoping, 2014). Li & Atick's

theory is motivated by the idea that binocular vision is organized to maximize the efficiency with which the two eyes' signals are encoded, in line with general theories of efficient coding in vision (Barlow, 1961; Barlow & Foldiak, 1988; Field, 1994; Simoncelli & Olhausen, 2001; Zhaoping, 2014). Li and Atick show how the efficiency can be maximized by applying separate gain controls to the binocular summation and differencing channels; in the case of a high signal to noise ratio, the gain applied to the channel should be inversely proportional to the signal strength in that channel.

If there really are "channels" for the binocular sum and difference, then it should be possible to selectively desensitize them through adaption using stimuli that selectively stimulate one or other of the channels. Previous studies have shown that this is indeed the case: The summation and differencing channels can be suppressed by adaptation, specifically prolonged viewing of correlated and anti-correlated random visual noise (May et al. 2011; May & Zhaoping, 2016, 2019). With binocularly correlated noise, the noise samples in the two eyes are identical, so if each noise sample is denoted by **N**, then the response of the summation channel to correlated noise will be **N**+**N**=2**N**, whereas for the differencing channel it will be **N**–**N**=0. On the other hand with binocularly anti-correlated noise, in which the corresponding parts of the pattern in the two eyes are of opposite luminance polarity, the two eyes' noise samples will be **N** and –**N**, summing to zero and differencing to 2**N**. These previous studies did not investigate depth perception: They examined the effects of binocular adaptation on the perception of ambiguous stimuli that delivered different signals to each channel. However, the adaptation predicted by Li and Atick's theory should have important consequences for depth perception, and this was investigated in the current study.

Suppose we have two signals as a function of space, one to the left eye $\mathbf{S_L}$ and one to the right eye $\mathbf{S_R}$. Their sum $\mathbf{S_+}$ and difference $\mathbf{S_-}$ are given by

$$\mathbf{S_+} = \mathbf{S_L} + \mathbf{S_R} \quad (1)$$

$$\mathbf{S_-} = \mathbf{S_L} - \mathbf{S_R} \quad (2)$$

The response of the neuron can be calculated in two ways. We can calculate it from $\mathbf{S_L}$ and $\mathbf{S_R}$:

$$\text{Response} = \mathbf{K_L}\cdot\mathbf{S_L} + \mathbf{K_R}\cdot\mathbf{S_R}, \quad (3)$$

where '$\cdot$' is the dot product. Alternatively, we can calculate its response from $\mathbf{S_+}$ and $\mathbf{S_-}$:

$$\text{Response} = \mathbf{K_+}\cdot\mathbf{S_+} + \mathbf{K_-}\cdot\mathbf{S_-}. \quad (4)$$

Equations 3 and 4 give two different ways of calculating the same neuron's response. Combining these two equations, we have

$$\mathbf{K_L}\cdot\mathbf{S_L} + \mathbf{K_R}\cdot\mathbf{S_R} = \mathbf{K_+}\cdot\mathbf{S_+} + \mathbf{K_-}\cdot\mathbf{S_-} \quad (5)$$

$$= \mathbf{K_+}\cdot(\mathbf{S_L} + \mathbf{S_R}) + \mathbf{K_-}\cdot(\mathbf{S_L} - \mathbf{S_R}) \quad (6)$$

$$= (\mathbf{K_+} + \mathbf{K_-})\cdot\mathbf{S_L} + (\mathbf{K_+} - \mathbf{K_-})\cdot\mathbf{S_R} \quad (7)$$

From equation 7, it is clear that

$$\mathbf{K_L} = \mathbf{K_+} + \mathbf{K_-} \quad (8)$$

$$\mathbf{K_R} = \mathbf{K_+} - \mathbf{K_-} \quad (9)$$

and hence

$$\mathbf{K_+} = 0.5 \times (\mathbf{K_L} + \mathbf{K_R}) \quad (10)$$

$$\mathbf{K_-} = 0.5 \times (\mathbf{K_L} - \mathbf{K_R}) \quad (11)$$

In the efficient coding theory of stereopsis, the summation and differencing channels can adapt independently. Let $\mathbf{K}$ and $\mathbf{K'}$ denote the receptive fields before and after adaptation, respectively; then adaptation controls the gains g_+ and g_- for $\mathbf{K_+}$ and $\mathbf{K_-}$ to make

$$\mathbf{K'}_{+} = g_{+} \times \mathbf{K}_{+} \tag{12}$$

$$\mathbf{K'}_{-} = g_{-} \times \mathbf{K}_{-}. \tag{13}$$

Then, analogous to equations 8 and 9, receptive fields for $\mathbf{S_L}$ and $\mathbf{S_R}$ after adaptation are

$$\mathbf{K'_L} = \mathbf{K'}_{+} + \mathbf{K'}_{-} \tag{14}$$

$$\mathbf{K'_R} = \mathbf{K'}_{+} - \mathbf{K'}_{-} \tag{15}$$

If we let $g_{+} = g_{-} = 1$, then $\mathbf{K'}_{+}$, $\mathbf{K'}_{-}$, $\mathbf{K'_L}$ and $\mathbf{K'_R}$ are equal to the unadapted receptive fields, $\mathbf{K}_{+}$, $\mathbf{K}_{-}$, $\mathbf{K_L}$ and $\mathbf{K_R}$, respectively. Henceforth, then, we use $\mathbf{K'}_{+}$, $\mathbf{K'}_{-}$, $\mathbf{K'_L}$ and $\mathbf{K'_R}$ to refer to the receptive fields both before and after adaptation, with the unadapted receptive fields corresponding to the special case of $g_{+} = g_{-} = 1$.

Equations 12 and 13 show that the adaptation proposed by Li and Atick's theory changes the amplitudes of $\mathbf{K'}_{+}$ and $\mathbf{K'}_{-}$ without changing their respective shapes; however, adaptation can change both the amplitudes *and* the shapes of $\mathbf{K'_L}$ and $\mathbf{K'_R}$, thereby changing the preferred disparity as well as the sensitivities of the neuron. If g_{+} is positive but g_{-} is zero, then $\mathbf{K'_L} = \mathbf{K'_R}$, and the neuron will be tuned to zero disparity. If g_{-} is positive and g_{+} is zero, the neuron will be tuned to a disparity close to the separation between the major peak and trough in the $\mathbf{K'}_{-}$ receptive field. For intermediate ratios of g_{-} and g_{+}, the neuron's preferred disparity will fall between these two extremes. This is illustrated in Figure 1.

Each column of Figure 1 shows a different adaptational state, from $g_{+} = 0$ on the left to $g_{-} = 0$ on the right. The central column corresponds to no adaptation, i.e. $g_{+} = g_{-}$. Figure 1 contains three panels, each for a different neuron tuned to a particular binocular disparity. In each panel, the lower row gives the $\mathbf{K_L}$ or $\mathbf{K'_L}$ (blue) and $\mathbf{K_R}$ or $\mathbf{K'_R}$ (red) receptive fields, and the top row gives the corresponding $\mathbf{K}_{+}$ or $\mathbf{K'}_{+}$ (magenta) and $\mathbf{K}_{-}$ or $\mathbf{K'}_{-}$ (green) receptive fields. The $\mathbf{K_L}$ and

$\mathbf{K_R}$ receptive fields in the central column (corresponding to no adaptation) are calculated using equations A1 and A3 from the Appendix. The $\mathbf{K}_+$ and $\mathbf{K}_-$ receptive fields for the same neuron (also in the central column, corresponding to no adaptation) are calculated from $\mathbf{K_L}$ and $\mathbf{K_R}$ using equations 10 and 11. The $\mathbf{K'}_+$ and $\mathbf{K'}_-$ receptive fields are calculated from $\mathbf{K}_+$ and $\mathbf{K}_-$ using equation 12 and 13 with different gain values for the different adaptation conditions (i.e. different columns of Figure 1). The $\mathbf{K'_L}$ and $\mathbf{K'_R}$ receptive fields in each non-central column are then calculated from the $\mathbf{K'}_+$ and $\mathbf{K'}_-$ receptive fields in that column using equations 14 and 15.

Figure 1 shows that there are two key effects of adjusting the gain ratio, g_+/g_-. The first is that, as g_+/g_- increases (as would be expected after viewing binocularly anticorrelated adaptation stimuli), the preferred binocular disparity of the neuron decreases. This can be seen for each neuron by comparing the left and right eye receptive fields along the lower row of each panel. If this were the only effect of binocular adaptation, then adaptation to anticorrelated stimuli would increase the perceived depth, because the neuron that was best tuned to the test stimulus post-adaptation would be one that responded best to larger binocular disparities in the unadapted state; conversely adaptation to correlated stimuli would reduce the perceived depth. However, these effects of adaptation on disparity tuning can be overcome by the second effect of adjusting the gain ratio, g_+/g_-: post-adaptation, the contrast sensitivity of each binocular neuron depends on its preferred disparity in the unadapted state. This can be seen by comparing the different neurons within a particular column of Figure 1. In the unadapted state (central column), all the neurons have equal contrast sensitivity (i.e. amplitude of the $\mathbf{K_L}$ and $\mathbf{K_R}$ receptive field profiles are the same for neurons with different preferred binocular disparity). When g_+/g_- increases (e.g. after adaptation to anticorrelated stimuli), this selectively boosts the sensitivity of neurons that are normally tuned to small disparities (this is because these neurons mainly contribute to the binocular

summation channel, and so are strongly boosted by any increase in gain of that channel); this would have the effect of reducing the perceived depth because, post-adaptation, the most strongly responding neurons are likely to be those that are normally tuned to small disparities. Conversely, when g_+/g_- reduces (e.g. after adaptation to correlated stimuli), this selectively boosts the sensitivity of neurons that are normally tuned to larger disparities, which would increase the perceived depth.

In summary, increasing the ratio g_+/g_- reduces a neuron's preferred binocular disparity, but also selectively boosts the contrast sensitivity of neurons tuned to small disparities in the unadapted state; decreasing the ratio g_+/g_- increases a neuron's preferred binocular disparity, but also selectively boosts the contrast sensitivity of neurons tuned to large disparities in the unadapted state. These effects would be expected to influence the perceived depth in opposite ways. In computational modelling reported later in this article, we show that the effects of adaptation on contrast sensitivity dominate: adaptation to binocularly anticorrelated stimuli boosts the sensitivity of neurons selective for small disparities, which reduces perceived depth; adaptation to binocularly correlated stimuli boosts the sensitivity of neurons selective for larger disparities, which increases perceived depth.

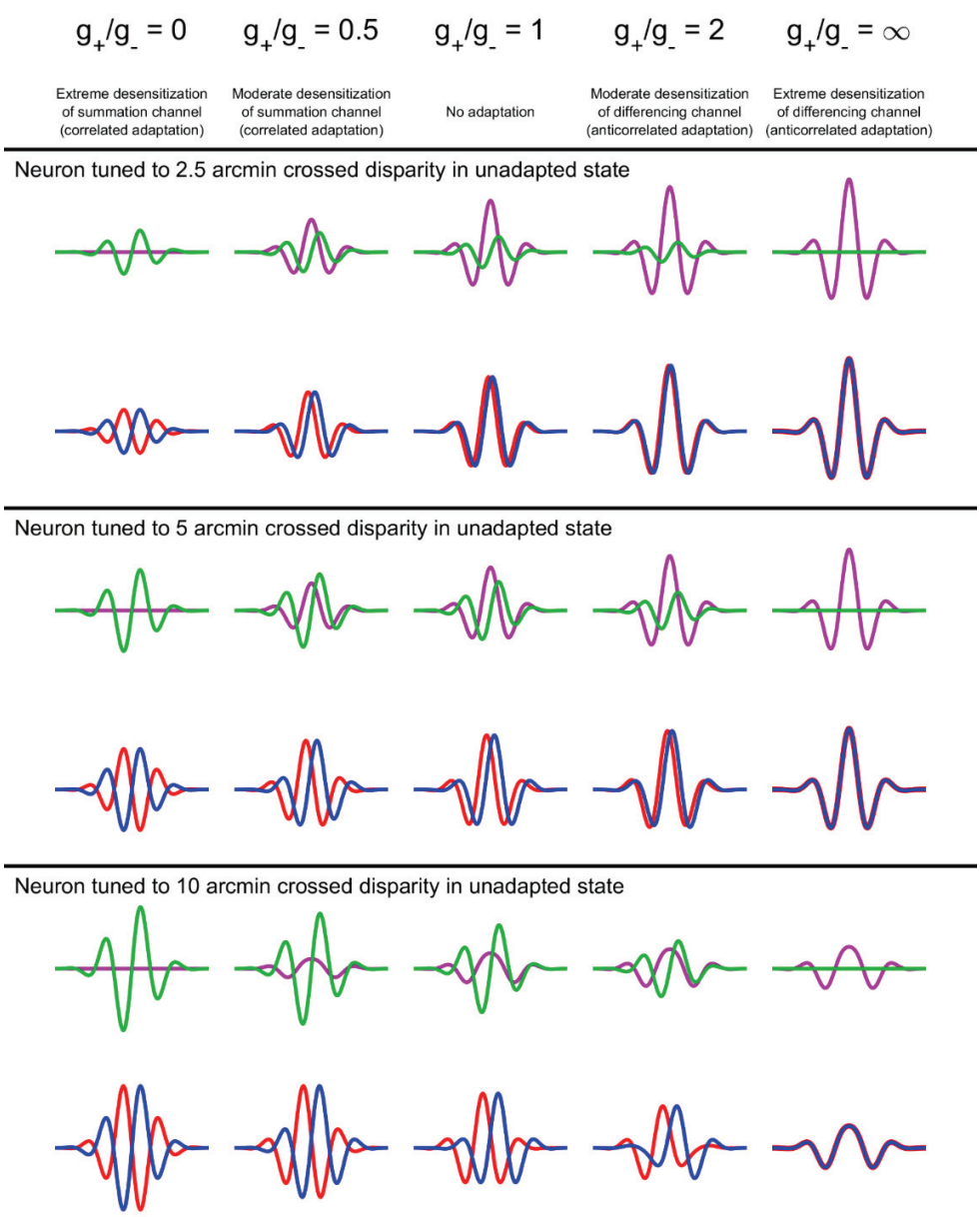

Figure 1. Effects of binocular adaptation on neural tuning properties. The three panels represent neurons tuned to different binocular disparities (all crossed disparities, i.e. target in front of the background). The five columns represent different adaptation conditions, with "no adaptation" in the middle. The lower row in each panel gives the neuron's receptive fields for the left eye (blue) and right eye (red). In the central column, these receptive fields are the unadapted receptive fields, $\mathbf{K}_L$ and $\mathbf{K}_R$, calculated using equations A1 and A3 from the Appendix, with the same parameters as used in the modelling that we report later in the article. The other columns of the lower row in each panel show the post-adaptation receptive fields, $\mathbf{K}'_L$ and $\mathbf{K}'_R$. The upper row in each panel gives the neuron's $\mathbf{K}_+$ or $\mathbf{K}'_+$ (magenta) and $\mathbf{K}_-$ or $\mathbf{K}'_-$ (green) receptive fields. See text for further details.

How the gains on the binocular channels can be independently adjusted is not specified in Li and Atick's theory, but it is presumably via suitable adjustments to the pattern of weights of the afferent inputs to the binocular neuron, for example LGN (lateral geniculate nucleus) or early stage V1 afferents. An important difference between the adaptation suggested by Li and Atick's theory and conventional adaptation models is that, in conventional adaptation models, the gain adjustment is a function of the activity of the neuron, whereas in Li and Atick's theory, the gain adjustment is determined by the interocular correlation, regardless of the neuron's activity level.

In the current study, we tested the predictions of Li and Atick's theory by measuring the comparative effects of adaptation to binocularly correlated versus anti-correlated adaptation stimuli on the perceived depth of a random-element stereogram. Examples of the adaptation and test stereograms are shown in Figures 2 and 7. Previous studies have shown that adaptation to a stereogram with a well-defined depth target affects the perceived depth of a similar depth target at a slightly different depth – this is the well-known 'depth after-effect' – and can be explained in terms of conventional adaptation models, which adjust the gain according to the activity of the neurons (Blakemore & Julesz, 1971; Long & Over, 1973; Rogers & Graham, 1985; reviewed by Howard & Rogers 1995). Our adaptation patterns consist of horizontally-oriented gratings and hence possess no horizontal stereo-disparities. Since binocular neurons tend to be tuned to horizontal disparities regardless of their preferred orientation (Cumming, 2002), the amount of stimulation that the binocular neurons receive from our horizontal adaptation stimuli would not depend on the neuron's preferred binocular disparity. Thus, a conventional adaptation model, in which adaptation is determined by the amount of stimulation a neuron receives, would predict no effect of our adaptation stimuli on the perceived depth of the test stimuli. In contrast, the efficient coding model of stereopsis predicts that perceived depth would be affected by our

adaptation stimuli because our correlated and anticorrelated adaptation conditions differ greatly in the relative responses of the binocular summation and differencing channels; this causes the two adaptation conditions to differ greatly in the relative gains applied to these channels, which would affect the perceived depth of the test stimulus as outlined earlier.

General Methods

Subjects

Six observers took part in the experiments. Two observers were authors, although at the time of testing one of them was naïve as to the purpose of the experiment. The remaining four were naïve adult volunteers. All observers had normal or corrected-to-normal visual acuity and good stereopsis as determined by the ability to reliably discriminate the relative depths of the test and match random-dot-stereogram targets used in the main experiment (a standard deviation of match settings less than 10% of the test disparity). Observers 1-4 provided data for the first experiment, Observers 3 and 4 for the second, and Observers 1, 3, 4, 5 and 6 for the third.

Prior to experimental testing informed consent was obtained from each observer. All experiments were conducted in accordance with the Declaration of Helsinki and the Research Institute of the McGill University Health Centre (RI-MUHC) Ethics Board. Observer initials on graphs have been anonymized in accordance with requirements of the (RI-MUHC) Ethics Board.

Apparatus

All experiments were conducted using a Dell Precision T1650 PC with a ViSaGe graphics card (Cambridge Research Systems (CRS), UK). The visual stimuli were displayed on a gamma-corrected Sony Trinitron Multiscan F500 flat-screen CRT Monitor. Stimulus generation and

experimental control employed C software. Participants viewed the dichoptic pairs through a custom-built 8-mirror Wheatstone stereoscope with an aperture of 10^o x 10^o and a viewing distance along the light path of 55 cm. During the experiments observers were seated in a darkened room and their responses were recorded via a keypad.

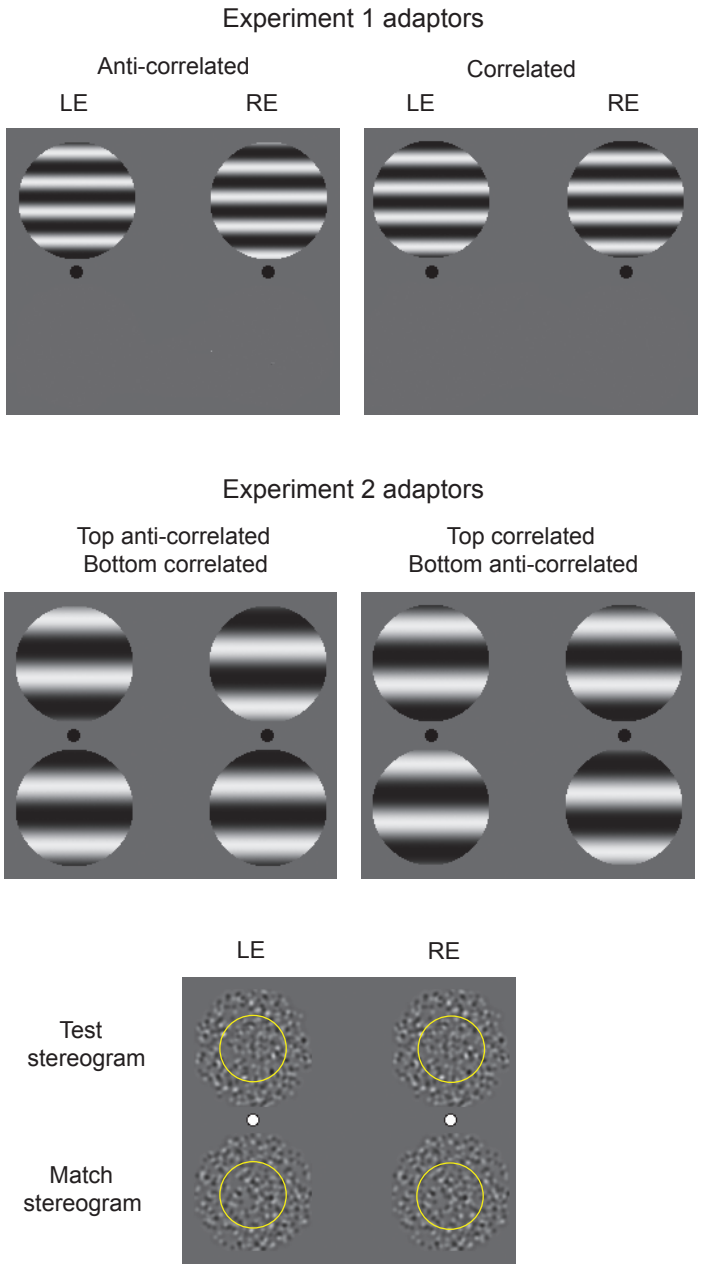

Figure 2. Experiments 1 and 2: example stimuli and protocols. In Experiment 1 only one adaptor was used per session, either anti-correlated (top left), correlated (top right) or none (not shown). In Experiment 2 both correlated and anti-correlated adaptors were presented together during a session, either anti-correlated on top and correlated on bottom (middle left), or vice-versa (middle right). For both experiments the test and match stereograms were always in the same spatial relationship as shown: test on top, match on bottom. The yellow circles indicate the position of the depth target, but were not present in the actual experiment

Stimuli

Experiments 1 and 2. Example stimuli for the first two experiments are shown in Fig. 2. The stimuli were circular patches with a diameter of 4.7 deg, positioned above and below fixation, separated vertically by a gap of 1.1 deg. All stimuli were spatial-frequency narrowband. The adaptors comprised single-spatial-frequency horizontally-oriented gratings with a contrast of 0.99. The absolute spatial phases of the adaptor gratings were randomized on each trial, but the relative phases of the left-eye and right-eye gratings were such as to produce the two main adaptor types: correlated, in which the noise textures were identical in each eye, and anti-correlated, in which they were of opposite luminance polarity in each eye. Unlike the adaptors, the test and match stimuli were orientationally broadband, constructed by summing randomly-positioned circularly-symmetric Laplacian-of-Gaussian (LoG) micropatterns, whose luminance profiles were given by:

$$LoG(r,\sigma) = \frac{-1}{\pi}\left[1 - \frac{r^2}{2\sigma^2}\right]\exp\left(\frac{-r^2}{2\sigma^2}\right) \qquad (9)$$

where r is radial distance and σ standard deviation. There were 3 values of σ, 0.05, 0.1 and 0.2 deg, and the number of micropatterns per stimulus was respectively 2000, 500 and 125, resulting in equal coverage. The root-mean-square (RMS) contrasts of the test and match stimuli was on average 0.195. Each test or match stimulus contained a depth target that stood in front of the background. This was achieved by offsetting the LoGs horizontally with sub-pixel accuracy in opposite directions in the two stereo-halves. The horizontal separation of the members of each

dichoptic pair on the monitor surface was adjusted so that the dichoptic pair appeared fused in the centre of the aperture of the stereoscope.

Experiment 3. Example stimuli are shown in Fig. 7. In this experiment, the test and match stereograms approximated fractal or 'pink' noise as such stimuli have the $1/f$ (f = Fourier frequency) amplitude spectra characteristic of images of natural scenes. They were constructed by summing randomly-positioned Gabors of various spatial frequencies and orientations, as described in detail elsewhere (Kingdom, Hayes & Field, 2001). Each Gabor had a spatial-frequency bandwidth of 1.59 octaves and odd-symmetric spatial phase. The set of Gabor spatial frequencies, f_G, were geometrically spaced at 0.51, 0.88, 1.53, 2.66 and 4.63 cpd. At each spatial frequency there were 12 orientations arranged at 30 deg. intervals. Each texture contained a total of 8100 Gabors, all with the same peak-to-trough amplitude and with the number at each spatial frequency proportional to the square of the spatial frequency. This results in a stimulus whose Fourier amplitudes are proportional to $1/f$ but with an equal amount of contrast energy in each octave Fourier frequency band. The overall RMS contrast of the Gabor textures was 0.214. As with the test and match stimuli in the first two experiments, the disparities of the depth targets were achieved by offsetting the Gabors in the target region horizontally with sub-pixel accuracy in opposite directions in the two stereo-halves. The adaptors were horizontal, multi-spatial-frequency gratings, composed of sine-wave components with the same spatial frequencies as the Gabor micropatterns used to construct the test and match stimuli, i.e. f = 0.51, 0.88, 1.53, 2.66 and 4.63, with amplitude proportional to $1/f$. The spatial phase of each spatial frequency component was randomized on each stimulus presentation. Dichoptic pairs of patterns, each 10.4 x 4.9 deg,

with one pair above and one below fixation, were separated vertically by a gap of 1.1 deg. The RMS contrast of the grating adaptors was 0.33.

Procedure

We employed the conventional sequence of adaptation followed by test followed by top-up adaptation followed by test etc., in conjunction with a staircase procedure that adjusted the disparity of the test targets according to previous responses. During the adaptation period, the stimuli were refreshed anew every 500ms. The adaptation sequence lasted 45 sec, i.e. there were 90 different presentations of the adaptor. Each test pattern was a single presentation preceded by and followed by an inter-stimulus-interval (ISI) of 100ms, interspersed with 6 secs of top-up-adaptation. The task on each trial was to identify which of the two target rectangles, the test or match, protruded more. The initial match disparity was randomly selected from a range that spanned 1.5 times the test disparity around its mean. A staircase method was employed to adjust the match disparity, such that a 'match in front of test' response resulted in a reduction in the match disparity and a 'match behind the test' response an increase. The staircase adjusted the match disparity by a factor of 1.4 for the first five trials and 1.1 from thereon. Thus the staircase converged on the PSE (point-of-subjective-equality) between the match and test. There were 20 test presentations per session and at the end of each session the geometric mean match disparity was calculated across the last 15 test trials. There were four to six sessions for each condition resulting in a total of 60 – 90 test trials per condition per observer.

Experiment 1. Effect of adaptation as a function of disparity

The stimulus protocol is illustrated in Figure 2. In this experiment there were three adaptor conditions: correlated, anti-correlated and no-adaptor (a blank). The adaptors were horizontally-oriented gratings of spatial-frequency 1.84 cpd that were always positioned above fixation. The test stereogram was constructed from LoG micropatterns with a σ of 0.2. In all cases the test stimulus was positioned above and the unadapted match stimulus below fixation.

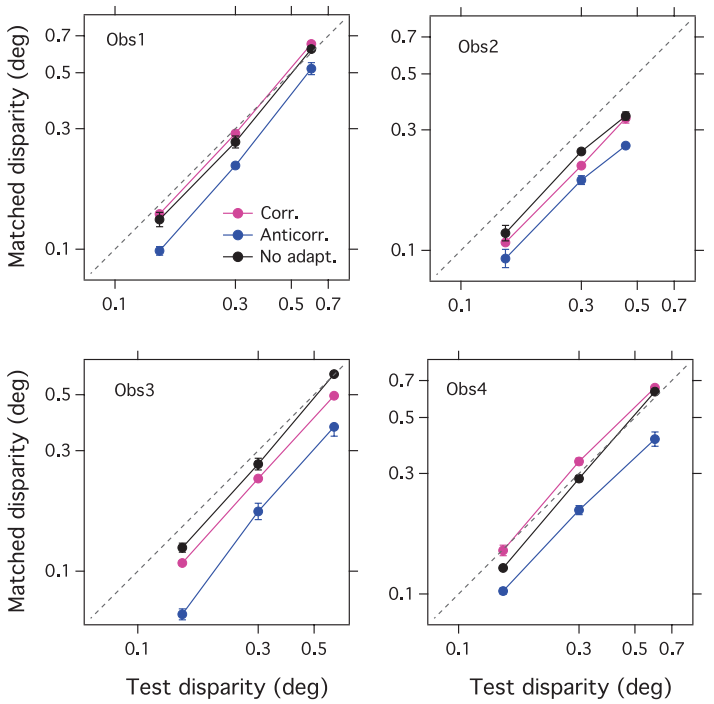

Figure 3. Results from Experiment 1, for four observers. The geometric mean match disparity at the PSE is plotted against test disparity, for three adaptor conditions: Corr. = correlated; Anticorr. = anti-correlated; No adapt. = no adaptor (blank). Error bars are upper and lower standard errors and are in some cases smaller than the symbols and hence not visible.

There were three test disparities: 0.15, 0.3 and 0.6 deg, except for observer 2 where the highest of the three disparities was set to 0.4 deg owing to the observer's difficulty in fusing the 0.6 deg disparity depth targets. Fig. 3 shows results for the four observers. The geometric mean disparity of the match stimulus at the PSE is shown as a function of the disparity of the test stimulus, for the three adaptor conditions. In all graphs the correlated adaptor data (magenta) lie above the anti-correlated adaptor data (blue), with no overlap in standard errors; thus, the target appeared to be further in front of the background after correlated than anticorrelated adaptation. The no adaptor data lie either above the correlated adaptor data or in between the correlated and anti-correlated adaptor data – we will return to a discussion of this finding later. Standard errors are in most cases smaller than the size of the data points (hence cannot be seen), demonstrating that observers performed the task with considerable precision, and that the differences between the two main adaptor types are very unlikely to be due to chance. Across observers and test disparities the geometric mean ratio (the antilog of the mean of the logarithm of the ratios) correlated-to-anticorrelated adaptor PSEs is 1.36, showing an average 36% shift in perceived depth between the two adaptors.

Experiment 2: **Effect of adaptation as a function of adaptor and test scale**

The protocol for the second experiment is illustrated in Fig. 2 and the set of stimuli employed in Fig. 4. As can be seen in Fig. 2, the difference from the first experiment is that both the correlated and anti-correlated adaptors were presented together, one above, the other below fixation. For each combination of adaptor spatial frequency and test/match micropattern σ, data were collected both with the correlated adaptor on top and the anti-correlated below – call this the C/A condition – and with the anti-correlated adaptor on top and the correlated below – call

this the A/C condition. However, the test was *always* above fixation and the match *always* below. The test disparity was 0.2 deg for all conditions. By comparing the C/A with A/C data without changing the relative positions of the test and match we are able to obtain a direct measure of the differential effect of the correlated and anti-correlated data without the confounding effects of visual field position.

Data for two observers are shown in Figure 5. Each graph shows C/A and A/C data as a function of the spatial frequency of the adaptors, with different graphs for the three test/match micropattern σs. As with the first experiment, in all conditions correlated adaptation produced greater perceived depth than anticorrelated adaptation (i.e. the test stimulus appeared to stand out further in front of the background after correlated than anticorrelated adaptation), with the difference reaching a maximum of about 60% in the 0.2 deg SD middle spatial frequency conditions.

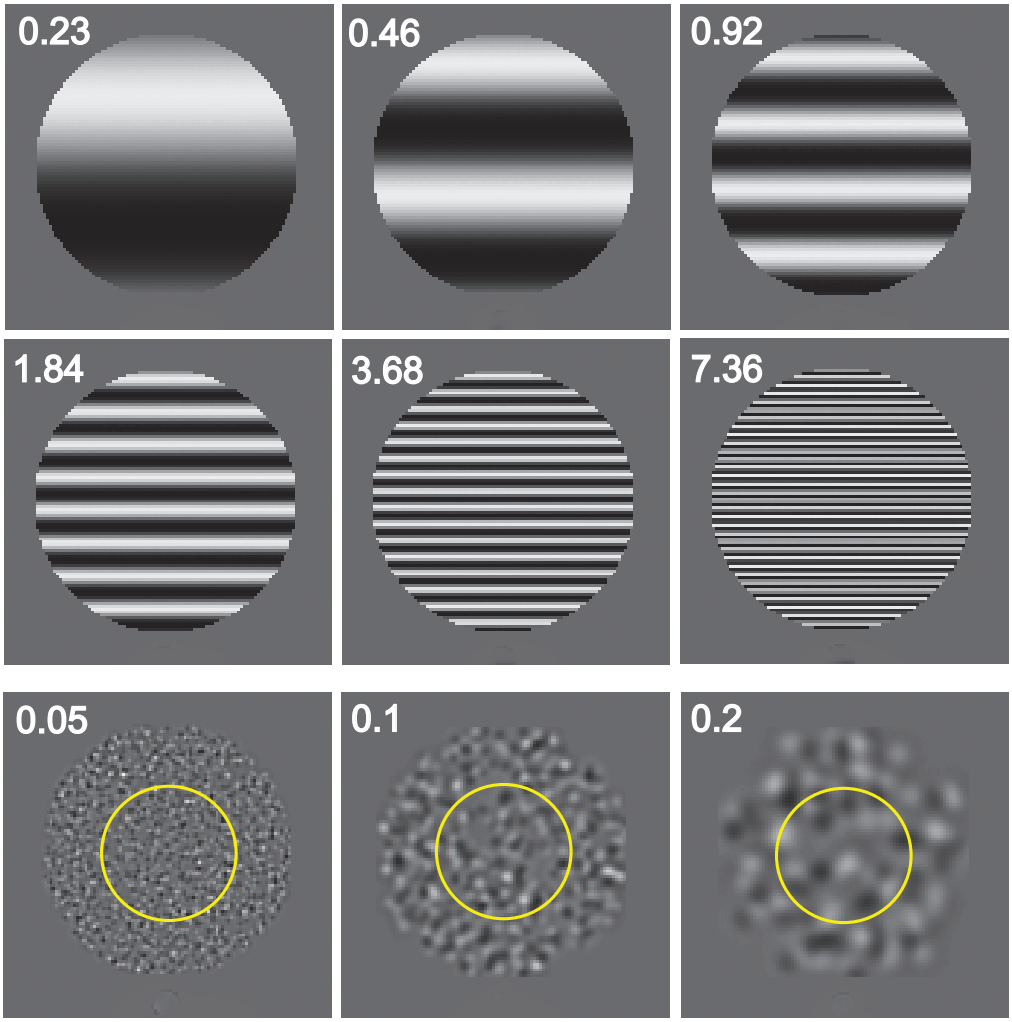

Figure 4. Experiment 2 stimulus conditions. Top two rows show the single sine-wave-grating adaptors, with spatial frequencies given in cpd. Bottom row shows the three tests, with SDs

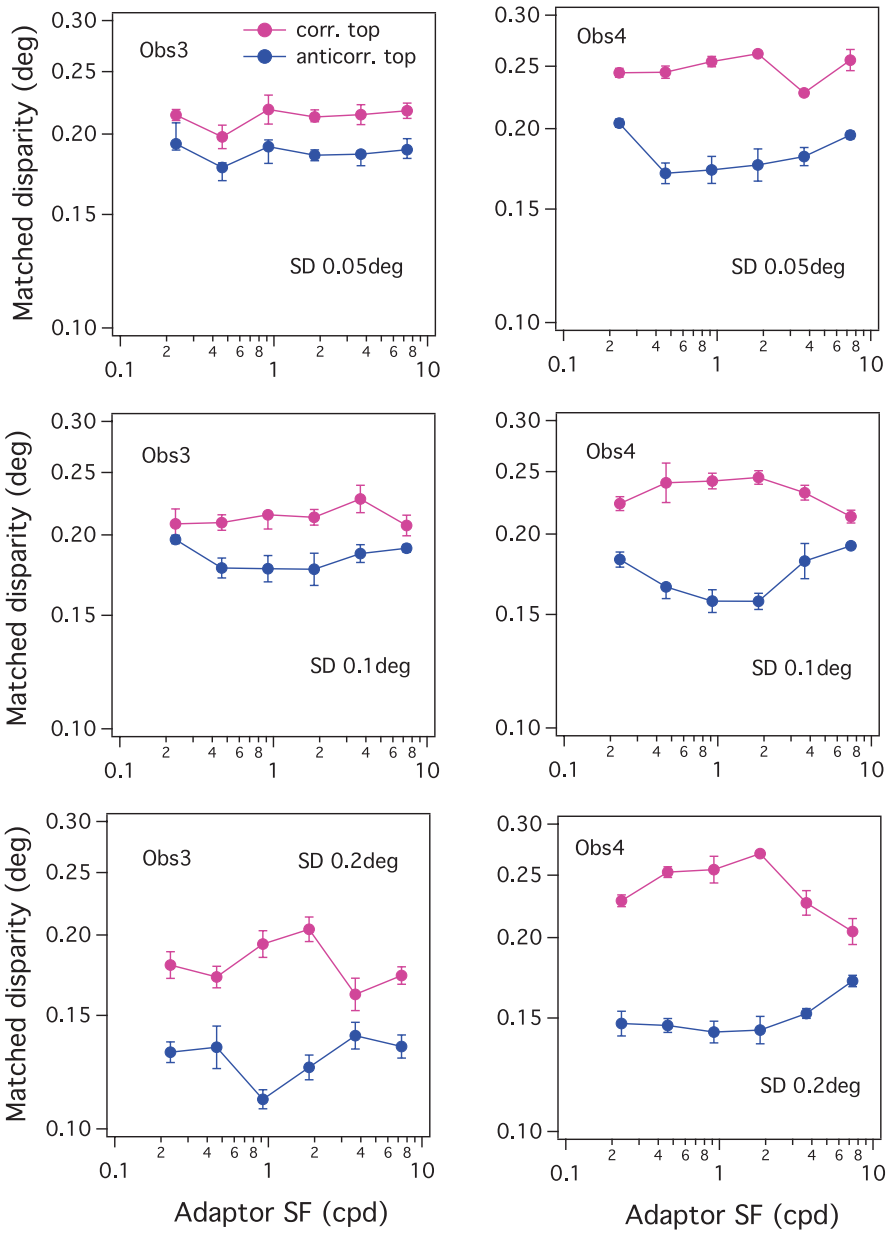

Figure 5. Experiment 2 results, for two observers. Geometric mean match disparity is plotted against adaptor spatial frequency, with the three test/match SDs in separate panels. Magenta points are for the C/A condition (correlated on top, anti-correlated on bottom), blue points for the reverse A/C conditions.

Fig. 6 shows the data from Experiment 2 expressed as the ratio of C/A to A/C PSEs. While the results are somewhat inconsistent across the two observers, the data show quite broad tuning for adaptor spatial frequency with a peak in most conditions around 1-2cpd, and with no hint of any difference in tuning depending on the test/match SD.

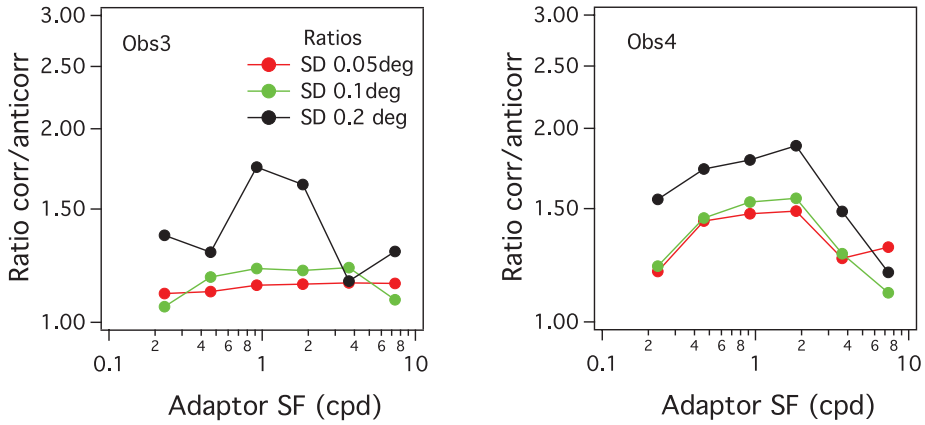

Figure 6. Experiment 2 results expressed as the ratio of C/A to A/C PSEs, for the three test/match SDs.

Across the two observers and adaptor SFs the mean geometric ratio of C/A to A/C PSEs is 1.32, showing an average 32% shift in perceived depth between the two adaption consitions, with a peak for the σ = 0.2deg condition across observers of around 80%.

Experiment 3. Broadband stimuli

The above experiments used narrowband-spatial-frequency adaptors and test stimuli, so the question arises as to whether our findings extend to broadband-spatial-frequency stimuli. To examine this we used test stimuli with the 1/*f* amplitude spectra characteristic of natural scenes, as shown in Fig. 7, with results for five observers shown in Fig. 8. The protocol was essentially the same as for Experiment 2. There were three test disparities: 0.1, 0.2 and 0.4 deg.

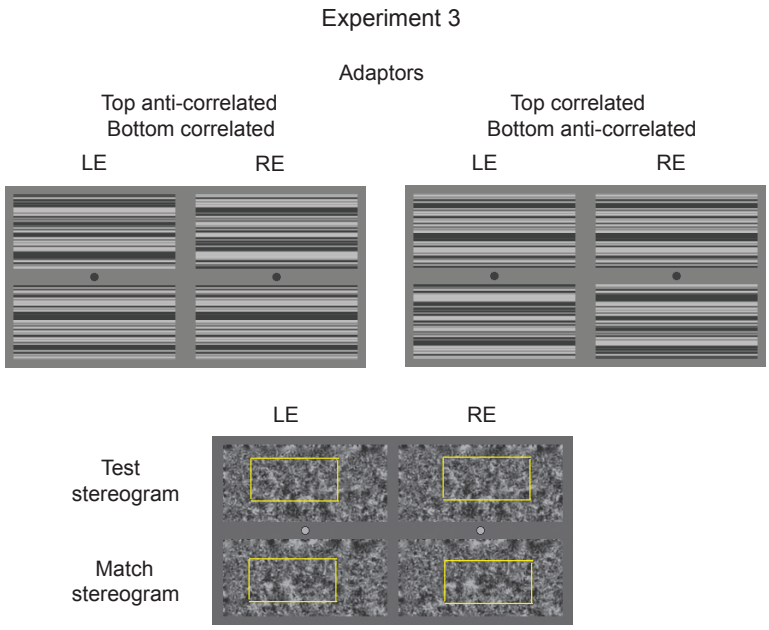

Figure 7. Experiment 3 example stimuli and protocol. See Methods for stimulus details.

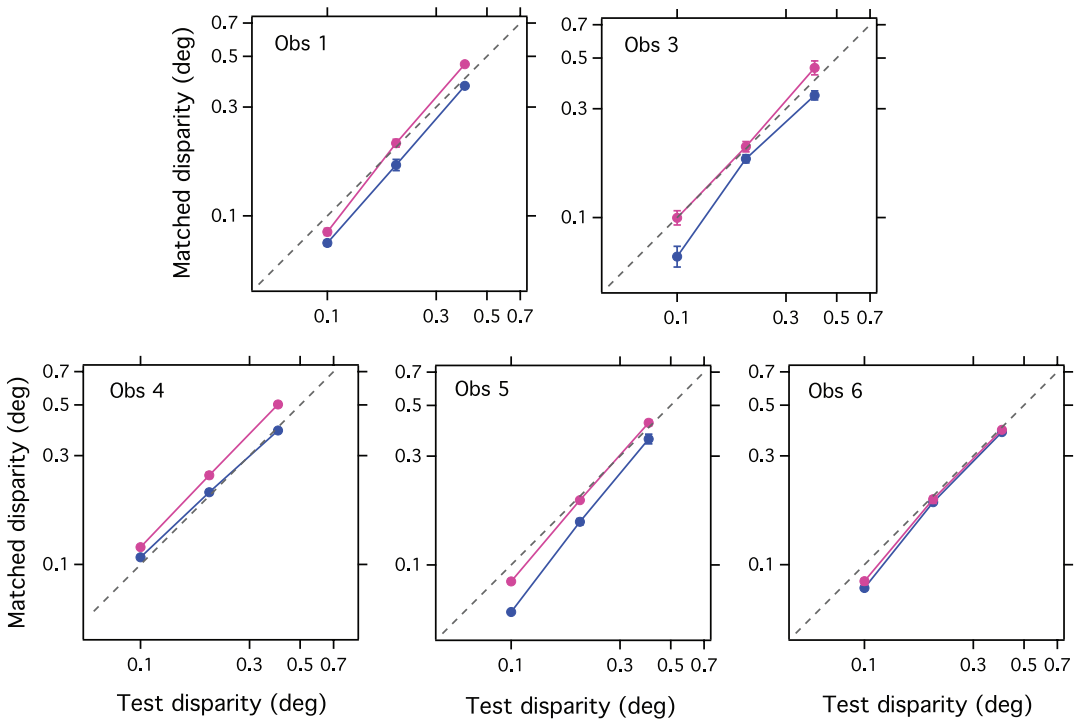

Figure 8. Experiment 3 results, for five observers. Magenta symbols: C/A; blue symbols A/C conditions. Error bars are upper and lower standard errors.

Across observers and test disparities the mean geometric ratio of C/A to A/C PSEs is 1.20, showing an average 20% shift in perceived depth between the two adaption conditions.

Discussion

Across three experiments we found that adaptation to horizontally-oriented grating patterns changed the apparent depth of a random-dot-stereogram target depending on whether the patterns were correlated or anti-correlated. The average differential shifts in perceived depth for

the three experiments were 36%, 32% and 20%, with the maximum differential shift being 80% for the 1-2 cpd adaptors and largest test SD (0.2) condition in Experiment 2.

In the first experiment we included a no-adaptation condition and depending on the observer found that the matched disparities either lay between those from the correlated and anti-correlated adaptor conditions or above those of the correlated adaptor condition. The adaptor in this experiment was presented only to the visual field to which the test was presented, and as a result there was a pronounced reduction in the perceived contrast of the test following adaptation. It is known that reducing the contrast of a low-spatial-frequency binocular target makes it appear further away (Rohaly & Wilson, 1999; Schor & Howarth, 1986). This shift in apparent depth of the test due to contrast adaptation would reduce the match disparity. It is possible that the reduction in match disparity due to contrast adaptation was larger than the increase in match disparity due adaptation to the interocular correlation, at least for some observers. This would explain why, for those observers, the match disparity for the correlated adaptation condition was smaller than for the no adaptation condition. This additional effect on perceived depth was not present in the other two experiments where both the test and match stimuli were subjected to the same amount of contrast adaptation.

Another finding of note is that in the second experiment there was no hint of an interaction between the spatial frequencies of the adaptor and the test in terms of the differential effect of correlated and anti-correlated adaptation on perceived depth. This finding runs counter to previous evidence for spatial-frequency tuned stereo-channels (reviewed by Howard & Rogers, 1995). Instead it is consistent with the idea that the adaptation in our experiments is mediated by a single spatial-frequency channel that responds best to 1-2 c/deg, with the most effective test a σ of 0.2deg: the Fourier transform of a Laplacian of Gaussian with σ of 0.2 deg

peaks at 1.13 cpd [this can be obtained from Cyganek & Siebert's (2009) Equation 4.79, replacing ω with $2\pi f$]. This reinforces the point that the adaptation effects described here have a different origin to those of conventional depth aftereffects.

In the next section, we show that the efficient coding theory predicts our results when incorporated into a standard model of population-coding model of binocular vision. Before that, we speculate why the adaptation mechanism that we identified in our experiments is fixed at a low spatial frequency (1-2 c/deg). The adaptation that we have described reduces the sensitivity of more strongly responding channels and boosts the sensitivity of less strongly responding channels, so that the channel responses are more equal. This is a form of whitening, and it can be shown that this maximises efficiency when the overall signal is strong. However, when the overall signal strength is low, it turns out that efficiency is improved by suppressing the weaker signal more than the stronger one, because a very weak (and therefore noisy) signal costs more in power consumption than is gained in information (Zhaoping, 2014; Li & Atick, 1994). Because of the 1/f spectrum of natural images, the overall signal is usually strong only for the low spatial frequencies. Thus, in normal viewing conditions, it is only in the low-spatial-frequency region where it is beneficial to whiten the signals. Thus we might expect that only the low-spatial-frequency channels would have a whitening mechanism that would reduce sensitivity in response to a strong signal. Fig 2A from Li & Atick (1994) shows that whitening of the binocular channels is only beneficial for spatial frequencies of around 1.5 c/deg or less. Thus, the reduced sensitivity of the binocular channels after strong stimulation during the adaptation phase would be expected only for channels tuned to spatial frequencies less than around 1.5 c/deg.

Modeling

We examined the predictions of a standard multi-channel model of binocular vision with an adaptation regime based on the efficient coding theory, in which the gain of each neuron's $\mathbf{K}_+$ or $\mathbf{K}_-$ receptive field is adjusted according to the interocular correlation, leading to a change in the neuron's preferred binocular disparity (see Appendix for details). The model contained an array of neurons tuned to different binocular disparities, and the model's "perceived" disparity was taken to be the preferred disparity (in the unadapted state) of the most strongly responding neuron. This approach makes use of both the strength of response of each model neuron, and its preferred disparity prior to adaptation, in estimating depth from the population response. The effect of adaptation on the gain of responses, as well as the change in disparity tuning, must be taken into account in determining perceived depth.

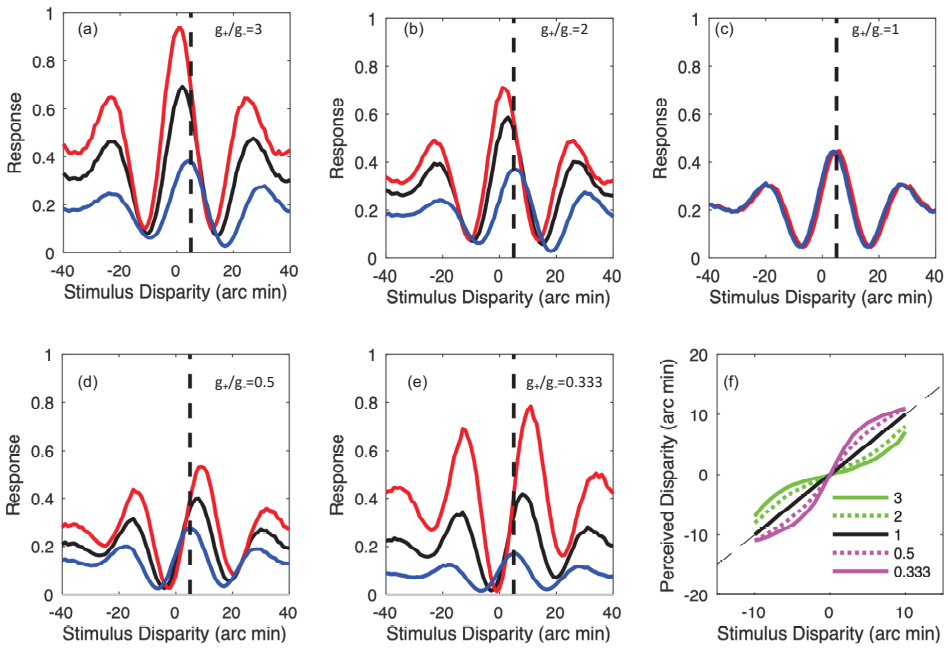

Figure 9. (a)-(e) show disparity tuning functions for three filters, following adaptation. In all cases, the black curve shows the response of a filter tuned to 5 arc min (the stimulus disparity is marked by the dashed vertical line). The red curve shows the tuning function that produced the maximu response at this stimulus disparity. The blue curve shows the response the response of another filter that has its peak at the stimulus disparity. The stimulus disparities of the latter two are indicated in the plots. (a) Adaptation to an anti-correlated stimulus, giving a gain ratio between summation and differencing channels of 3. (b) Adaptation to an anti-correlated stimulus, giving a gain ratio between summation and differencing channels of 2. (c) Adaptation to an uncorrelated stimulus, giving a gain ratio between summation and differencing channels of 1. (d) Adaptation to a correlated stimulus, giving a gain ratio between summation and differencing channels of 0.5 (e) Adaptation to a correlated stimulus, giving a gain ratio between summation and differencing channels of 0.33. (f) The estimated disparity, taken as the pre-adaptation preferred disparity of the filter with the strongest response, as a function of stimulus disparity for the 5 adaptation states shown.

In the Introduction, we showed that anticorrelated adaptation would tend to reduce the neuron's preferred disparity, while selectively enhancing the sensitivity of neurons normally tuned to small disparities; conversely, correlated adaptation would tend to increase the neuron's preferred disparity, while selectively enhancing the sensitivity of neurons normally tuned to larger disparities. The implication for these changes in disparity tuning function shape and amplitude are illustrated in figure 9. Panels (a) to (e) of this figure each show disparity tuning functions for three model neurons, following a particular adaptational state. (a) and (b) illustrate the case in which $g_+/g_- > 1$, as would be expected after anticorrelated adaptation; (c) illustrates the case in which $g_+/g_- = 1$, as would be expected after no adaptation; (d) and (e) illustrate the case in which $g_+/g_- < 1$, as would be expected after correlated adaptation. In each panel, the absolute gains (and hence absolute responses) are not important, as the model's perceived depth is a function of the ratio g_+/g_-, not the absolute sensitivities. The black curve in each panel gives the tuning function of the neuron that is tuned to a disparity of −5 arcmin (i.e. 5 arcmin crossed disparity) in the unadapted state; this disparity is indicated by the vertical dashed line). As expected, the peak response of this neuron is to a stimulus with disparity closer to zero after anticorrelated adaptation, and further from zero after correlated adaptation. In each panel (a) to (e), the neuron that is best tuned to a disparity of −5 arcmin after adaptation (shown in blue) is not the neuron with the strongest response. The neuron with the strongest response to the −5 arcmin disparity in each adaptation condition is shown in red. The model's estimated disparity is the preferred disparity of this neuron in the unadapted state. Because of the predicted effects of binocular adaptation on the contrast sensitivities of different neurons explained earlier, the most strongly

responding neuron in the anticorrelated conditions has a pre-adaptation preferred disparity smaller than the stimulus disparity, and the most strongly responding neuron in the correlated conditions has a pre-adaptation preferred disparity larger than the stimulus disparity. This accounts for the psychophysical results that we observed. We cannot quantitatively predict the gain ratio g_+/g_- after adaptation, so we have derived the model's performance for a range of ratios, as shown in Figure 9(f). For $g_+/g_- > 1$ (anticorrelated adaptation, shown in blue), the estimated disparity is closer to zero than the true disparity; for $g_+/g_- < 1$ (correlated adaptation, shown in red), the estimated disparity is further from zero than the true disparity. Thus, the efficient coding model predicts large changes in preferred disparity in the direction obtained in the experiment (the absolute size of the change in disparity is not important because this will vary depending on the exact parameters of the model, which are not constrained by our experiment).

It should be noted that in our model, the effects of adaptation on the neuronal response is determined by the interocular correlation of the adaptation stimuli. This is radically different from a conventional adaptation model, in which the effects of adaptation are determined by the neuronal activity during the adaptation period. Our adaptation stimuli were designed so that the amount that they stimulated the neurons would be unaffected by the neuron's preferred disparity, and thus a conventional model of adaptation would predict that perceived depth would be unbiased by our adaptation protocol. The large effects of adaptation on perceived depth that we obtained therefore provide strong support for a more sophisticated type of adaptation, in which the response characteristics of binocular neurons are adjusted according to the prevailing interocular correlation, and not just the activity level of the neurons, as suggested by Li and Atick's theory of efficient binocular coding.

Conclusion

We have shown that adaptors with no horizontal disparities can induce significant differential effects on the perceived depth of a random-dot-stereogram target. This finding supports Li and Atick's efficient coding theory of stereopsis in which binocular summation and differencing channels are multiplexed within a binocular neuron, and the responses of these channels are separately adaptable, with gain changes being determined by the prevailing interocular correlation.

Appendix A

Multi-channel model details

The model consisted of a set of neurons tuned to different binocular disparities, defined according to the standard binocular energy model. The left-eye and right-eye receptive fields in the unadapted state are given by

$$K_{\mathrm{L,even}}(x) = e^{\left(\frac{-x^2}{2\sigma^2}\right)}\cos\,(2\pi f x) \quad \text{(A1)}$$

$$K_{\mathrm{L,odd}}(x) = e^{\left(\frac{-x^2}{2\sigma^2}\right)}\sin\,(2\pi f x) \quad \text{(A2)}$$

$$K_{\mathrm{R,even}}(x;d) = e^{\left(\frac{-(x-d)^2}{2\sigma^2}\right)}\cos\,(2\pi f(x-d)) \quad \text{(A3)}$$

$$K_{\mathrm{R,odd}}(x;d) = e^{\left(\frac{-(x-d)^2}{2\sigma^2}\right)}\sin\,(2\pi f(x-d)) \quad \text{(A4)}$$

where d is the preferred binocular disparity, f is the spatial frequency tuning (2.4 cycles/degree) and $\sigma = 12.5$ arc min. For each neuron with unadapted preferred disparity d, post-adaptation summation and differencing receptive fields, even and odd versions of $K'_+(x)$ and $K'_-(x)$, are

given by Equations 5 and 6, with $K_{L,even}(x)$, $K_{L,odd}(x)$, $K_{R,even}(x)$ and $K_{R,odd}(x)$, given by equations A1-A4, and the gains, g_+ and g_- given by

$g_+ = 0.5 - k$

$g_- = 0.5 + k$

where $k = 0$ for no adaptation, $k = 1/4$ or $1/6$ for correlated adaptors and $k = -1/4$ or $-1/6$ for anticorrelated adaptors. The absolute magnitudes of g_+ and g_- do not matter; what matters is the ratio g_+/g_- after adaptation, rather than the specific values of g_+ and g_-: multiplying g_+ and g_- by the same constant (so the ratio is unchanged) will multiply all the neuronal outputs by a constant value, so the model's perceived depth will be unchanged. k values of 1/4, 1/6, 0, −1/6 and −1/4 give g_+/g_- ratios of 1/3, 1/2, 1, 2 and 3, respectively. The post-adaptation left- and right-eye receptive fields, $K'_L(x)$ and $K'_R(x)$, are then given by equations 7 and 8. Post-adaptation, the response to the test stimulus is given by

$$B_{post}(d) = \left[\left(\mathbf{K'_{L,even}} \cdot \mathbf{S_L} + \mathbf{K'_{R,deven}} \cdot \mathbf{S_R}\right)^2 + \left(\mathbf{K'_{L,odd}} \cdot \mathbf{S_L} + \mathbf{K'_{R,odd}} \cdot \mathbf{S_R}\right)^2\right] \qquad \text{(A5)}$$

where $\mathbf{S_L}$ and $\mathbf{S_R}$ are the left- and right-eye test signals.

Calculating model response. The test stimulus is always correlated pink Gaussian noise, with a varying disparity. For each disparity, 1000 samples of $B_{post}(d)$ were generated, the mean taken over these samples, then the expected depth estimate taken as the peak of this function. The results plotted in Fig. 9 are for a single spatial frequency. Since the noise is 1/f (so looks the same at any scale), changing the size of the filter kernel only has the effect of linearly scaling the graphs along the horizontal and vertical axes.

Acknowledgments

This work was funded by the Canadian Institute of Health Research grant #MOP 123349 given to F.K. (Commercial relationships: none).

References

Anzai A., Ohzawa I. & Freeman R. D. (1999). Neural mechanisms for encoding binocular disparity: Receptive field position vs. phase. *Journal of Neurophysiology,* 82 (2), 874–890.

Atick, J. J., Li, Z. & Redlich, N. (1993), What does post-adaptation color appearance reveal about cortical color representation? *Vision Research*, 33, 123-129.

Barlow, H. (1961) Possible principles underlying the transformation of sensory messages in *Sensory Communication*, MIT Press.

Barlow, H. B. & Foldiak, P. (1989.) Adaptation and decorrelation in the cortex. In *The Computing Neuron,* ed. C. Miall, R. M. Durbin and G. J. Mitchison, pp. 54–72. Wokingham, UK: Addison-Wesley

Blakemore, C. & Julesz, B. (1971). Stereoscopic depth aftereffect produced without monocular cues. *Science*, 171, 286-288.

Buchsbaum, G. & Gottschalk, A. (1983). Trichromacy, opponent colour coding and optimum colour information transmission in the retina. *Proc. R. Soc. Lond. B Biol. Sci*. 220:89–113

Cohn, T. E., & Lasley, D. J. (1976). Binocular vision: Two possible central interactions between signals from two eyes. Science, 192, 561–563.

Cohn, T. E., Leong, H. & Lasley, D. J. (1981). Binocular luminance detection: availability of more than one central interaction. *Vision Research*, 21, 1017-1023.

Cumming, B. G. (2002). An unexpected specialization for horizontal disparity in primate primary visual cortex. *Nature*, 418, 633–636.

Cumming, B. G. & Parker, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature*, 389, 280–283

Cyganek, B., & Siebert, J. P. (2009). *An Introduction to 3D Computer Vision Techniques and Algorithms*. Chippenham: John Wiley & Sons, Ltd.

DeSilva, H. R., & Bartley, S. H. (1930). Summation and subtraction of brightness in binocular perception. *British Journal of Psychology*, 20, 242–252.

Field, D. J. (1994). What is the Goal of Sensory Coding? *Neural Computation*, 6, 559-601.

Georgeson, M. A, Wallis, S. A., Meese, T. S. & Baker, D. H. (2016). Contrast and Lustre: a model that accounts for eleven different forms of contrast discrimination in binocular vision. *Vision Research* 129, 98-118.

Henriksen, S. & Read, J. C. A. (2016). Visual Perception: A Novel Difference Channel in Binocular Vision. *Current Biology*, **26**(12), R500-R503.

Howard, I. P. & & Rogers, B. J. (1995). *Binocular Vision and Stereopsis*. Oxford University Press, New York.

Kingdom, F. A. A., Hayes, A. & Field, D. J. (2001) Sensitivity to contrast histogram differences in synthetic wavelet-textures. *Vision Research*, 41, 585-598.

Li, Z. & Atick, J. J. (1994). Efficient stereo coding in the multiscale representation. *Network: Computation in Neural Systems*, 5, 157-174.

Long, N. R. & Over, R. (1973). Stereoscopic depth aftereffects with random dot patterns. *Vision Research*, 13, 1283-1287.

Kingdom, F. A. A. (2012). Binocular vision: The eyes add and subtract. *Current Biology*, 22, R22-24.

May, K., Zhaoping, L. & Hibbard, P. (2012). Perceived direction of motion determined by adaptation to static binocular images. *Current Biology*, 22, 28-32.

May, K. A. & Zhaoping, L. (2016). Efficient coding theory predicts a tilt aftereffect from viewing untilted patterns. *Current Biology*, 26, 1571-1576.

May, K. A., & Zhaoping, L. (2019). Face perception inherits low-level binocular adaptation. *Journal of Vision, 19*(7):7, 1-10.

Ohzawa I., DeAngelis, G. C. & Freeman R. D. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science,* 249*,* 1037–1041.

Read, J. C., & Cumming, B. G. (2004). Understanding the cortical specialization for horizontal disparity. *Neural computation*, *16*(10), 1983-2020.

Rogers, B. J. & Graham, M. E. (1985). Motion parallax and the perception of three-dimensional surfaces. In *Brain mechanisms and spatial vision* (ed. D. Ingle, M. Jeannerod and D. Lee) pp. 95-111. Martinus Nijhoff, The Hague.

Rohaly, A. M., & Wilson, H. R. (1999). The effects of contrast on perceived depth and depth discrimination. Vision Research, 39, 9-18.

Ruderman, D. L., Cronin, T. W. & Chiao, C-C. (1998). Statistics of cone responses to natural images: implications for visual coding. *J. Opt. Soc. Am. A* 15:2036–45.

Said, C. P. & Heeger, D. J. (2013). A model of binocular rivary and cross-orientation suppression. *PLoS Computational Biology*, 9(3): e1002991.

Schor, C. M., & Howarth, P. A. (1986). Suprathreshold stereo-depth matches as a function of contrast and spatial frequency. Perception, 15, 249-258.

Shevell, S. K. & Kingdom, F. A. A. (2008). Color in complex scenes. *Annual Review of Psychology*, 59, 143-166

Simoncelli, E. P. & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annuual Review Neurosci*ence, 24:1193–216.

Yoonessi, A., & Kingdom, F. A. A. (2009). Dichoptic difference thresholds for uniform color changes applied to natural scenes. *Journal of Vision*, 9(2), 1–12.

Zhaoping, Li (2014) *Understanding Vision: theory, models, and data*. Oxford University Press.