PERGAMON

Vision Research 39 (1999) 2313–2324

Vision Research

# Stereopsis from contrast envelopes☆

Keith Langley [a,*], David J. Fleet [b], Paul B. Hibbard [a]

[a] *Department of Psychology, Institute of Cognitive Neuroscience, University College London, London, WCIE 6BT, UK*
[b] *Department of Computing and Information Science, Queens University, Kingston, K7L 3N6, Canada*

Received 18 May 1998; received in revised form 30 September 1998

**Abstract**

We report two experiments concerning the site of the principal nonlinearity in second-order stereopsis. The first exploits the asymmetry in perceiving transparency with second-order stimuli found by Langley et al. (1998) (Proceedings of the Royal Society of London B, 265, 1837–1845) i.e. the product of a positive-valued contrast envelope and a mean-zero carrier grating can be seen transparently only when the disparities are consistent with the envelope appearing in front of the carrier. We measured the energy at the envelope frequencies that must be added in order to negate this asymmetry. We report that this amplitude can be predicted from the envelope sidebands and not from the magnitude of compressive pre-cortical nonlinearities measured by other researchers. In the second experiment, contrast threshold elevations were measured for the discrimination of envelope disparities following adaptation to sinusoidal gratings. It is reported that perception of the envelope's depth was affected most when the adapting grating was similar (in orientation and frequency) to the carrier, rather than to the contrast envelope. These results suggest that the principal nonlinearity in second-order stereopsis is cortical, occurring after orientation- and frequency-selective linear filtering.


*Keywords:* Second-order stereopsis; Multiplicative transparency; Depth asymmetry

## 1. Introduction

A central issue in stereopsis concerns the way in which left and right eyes views are combined to solve the correspondence problem and to recover three dimensional scene structure. Conventional models posit that matching relies on interocular correlation of monocular band-pass signals (e.g. Poggio & Poggio, 1984; Ohzawa, DeAngelis & Freeman, 1990; Blake & Wilson, 1991; Cormack, Stevenson & Schor, 1991; DeAngelis, Ohzawa & Freeman, 1995; Mallot, Arndt & Bulthoff, 1995; Fleet, Wagner & Heeger, 1996). But there are classes of signals (often called non-Fourier or second-order), the perception of which is not well captured by these models. Examples include texture boundaries (Frisby & Mayhew, 1978), motion boundaries (Halpern, 1991), and contrast envelopes (Liu, Schor & Ramachandran, 1992; Sato & Nishida, 1993; Hess & Wilcox, 1994; Fleet & Langley, 1994b; Sato & Nishida, 1994; Hibbard, Langley & Fleet, 1995; Lin & Wilson, 1995; Wilcox & Hess, 1995, 1996, 1997).

There are several plausible models for second-order visual processing, all of which involve an important nonlinearity at some stage. There are single-channel models that involve an early, pre-cortical nonlinearity (e.g. Burton, 1973). Such models have received attention in the context of second-order motion (e.g. Brown, 1995; Scott-Samuel & Georgeson, 1995). According to this model, the nonlinearity introduces a distortion product into the visual signal at the frequencies of the contrast envelope (Derrington & Badcock, 1986). This distortion product would then be processed by a conventional first-order model of binocular matching to deliver the disparity information present in the second-order signal.

There is increasing support, however, for a two-channel model. The two-channel models used to explain stereopsis resemble those forwarded in visual motion, with a separate channel for processing the second-order signal (Chubb & Sperling, 1988; Turano & Pantle, 1989; Victor & Conte, 1992; Wilson, Ferrera & Yo,

☆ Portions of this research were presented at ECVP 1994 and 1995 and Hibbard's Ph.D Thesis.

* Corresponding author. Fax: + 171-436-4276.
*E-mail address:* kl@psychol.ucl.ac.uk (K. Langley)

0042-6989/99/$ - see front matter 
PII: S0042-6989(98)00271-5

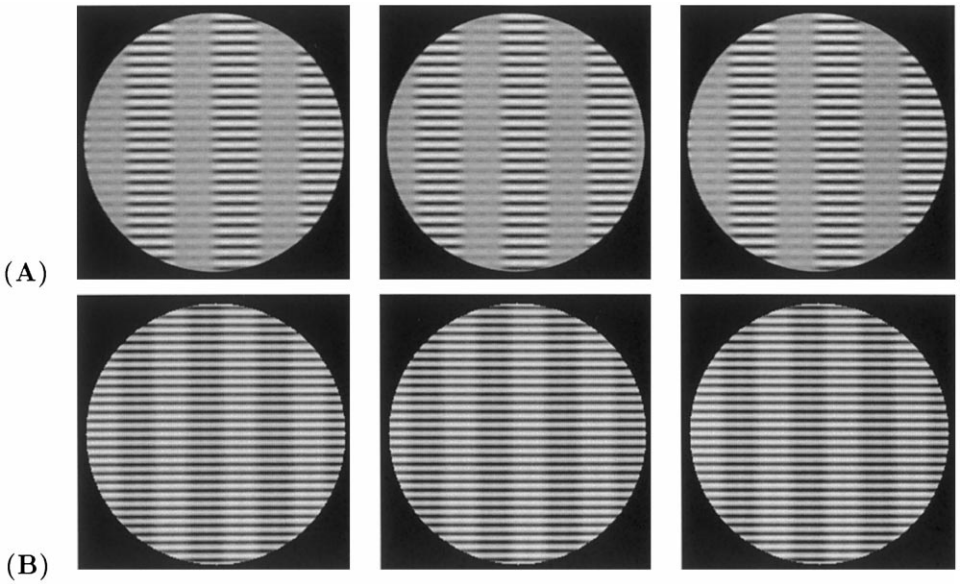


Fig. 1. Examples of the binocular stimuli. (A) A multiplicative combination of a vertically oriented contrast envelope and horizontal carrier grating. (B) Each image shows an additive combination of two gratings like those in (A). Cross-eyed fusion of the image pairs yields a transparency behind (left-center pair), or in front (center-right pair) of the horizontal grating.

1992; Zanker, 1993; Fleet & Langley, 1994a; Langley, 1997, 1999). These models differ in several ways, including both the site and the mathematical form of the nonlinearity. Many models suppose a cortical nonlinearity, so that the visual signal is processed by linear filters tuned to spatial frequency and orientation before the nonlinearity, and before second-order matching takes place.

The form of linear filtering that precedes the nonlinearity is informative experimentally. If one finds that this filtering is selective to spatial frequency and orientation, then a cortical nonlinerity is implied. Conversely, if it is broad-band and isotropic then a pre-cortical nonlinearity is implied, perhaps owing to the compressive nonlinearity of neurons found in the LGN (Sclar, Maunsell & Lennie, 1990; Scott-Samuel & Georgeson, 1995). Similar issues exist concerning the perception of monocular contrast envelopes (Burton, 1973; Henning, Hertz & Broadbent 1975; Derrington & Badcock, 1986; Langley, Fleet & Hibbard, 1996; Mareschal & Baker, 1998), and in the context of second-order motion (Chubb & Sperling, 1988; Wilson et al., 1992; Fleet & Langley, 1994a).

Wilcox and Hess (1996) posited that the nonlinearity occurs before the binocular combination of monocular signals. They also showed that second-order stereopsis could be difficult when the carrier orientations in contrast-modulated inputs where different, suggesting that orientation-specific processing precedes the nonlinearity. In contrast, Langley, Fleet and Hibbard (1998) (see also Liu et al., 1992) have shown that binocular depth perception from contrast envelopes is possible when carrier gratings differ significantly in spatial frequency. The reasons for this discrepancy could reflect differences in the paradigms and stimuli used.

This paper addresses the site of the nonlinearity with two experiments. The first exploits the results of Langley et al. (1998) who examined disparity thresholds for sums (first-order) and products (second-order) of binocular gratings like those used here (see Fig. 1). They found that, when viewed binocularly, the sum of two 1-d gratings of different spatial frequency and orientation may be seen transparently. Either grating can be perceived in front of or behind the other, depending on the disparities of the respective gratings. When the same two gratings were multiplied together transparency also occurred, but there was an asymmetry. The contrast envelope was only perceived in a separate depth plane when in front of the carrier; it was never seen as the underlaying surface.

To explain the depth asymmetry Langley et al. (1998) (see also Kersten, 1991) applied the constraints on the perception of monocular transparency proposed by Metelli (1974): (i) no matter how a monocular transparency is produced, the overlaying transparent surface must not change the order of luminance values reflected from the underlaying surface; and (ii) when lightness values are attenuated by a transparent surface, local differences in lightness seen through the transparent surface must be less than those seen without the transparent surface. If one assumes that these constraints influence binocular depth perception, then the asymmetry can be explained.

Metelli's constraints also predict that the product of the two positive-valued signals used in our study may be perceived symmetrically in depth; either signal may

be seen in-front of or behind the other. This is because both signals satisfy Metelli's monocular constraints on transparency so that a monocular cue could not be used by the visual system to influence binocular depth perception (see also Beck, 1984; Beck, Prazdny & Ivry, 1984; Kersten, 1991). This case, consistent with the results of Langley et al. (1998), is exploited in the first experiment here. In particular, one can add power at the envelope frequencies to a second-order signal to create a stimulus that is equal to a product of positive-valued gratings, and is therefore perceived symmetrically transparent. Here, we measured the amplitude of additive signal at envelope frequencies that was required to override the depth asymmetry. We then compared these thresholds to those predicted from two models, namely, a single channel model with a compressive nonlinearity as measured in other studies (Brown, 1995; Scott-Samuel & Georgeson, 1995), and a model that depends on the energy of the second-order signal. We report that this amplitude may be predicted from the sum of the contrast envelopes sidebands[1] rather than that expected from the magnitude of precortical nonlinearities as measured in other studies (Burton, 1973; Henning et al., 1975; Sclar et al., 1990; Scott-Samuel & Georgeson, 1995).

In our second experiment, subjects were adapted to a high contrast sinusoidal grating and then asked to detect the relative depth of the contrast envelope in a contrast-modulated test stimulus. If the site of adaptation was after the nonlinearity, then one would expect the most effective adapting frequencies to be similar to those of the envelope. If the site of the adaptation was before the nonlinearity, then we would expect that the carrier frequency would make the most effective adapting stimulus. If the effect of the adaptation was strongly orientation- and frequency-specific, then this implies that it occurs in visual cortex. We report that adaptation caused the largest threshold elevations when the frequency and orientation of the adapting grating were close to the carrier rather than to the envelope. Moreover, the adaptation was orientation and frequency specific. From these data, we conclude that the site of the nonlinearity occurs in the cortex, after orientation and frequency selective filtering.

## 2. Methods

### 2.1. Subjects

Three subjects were used for each experiment. Two subjects were authors. The third subject did not know the purpose of the experiments. All subjects had normal or corrected-to-normal vision.

### 2.2. Apparatus

Binocular images were presented on a SONY monitor with a refresh rate of 76 Hz and 256 grey scales. Eight-bit quantization, although crude, was thought sufficient because contrast thresholds for the discrimination of envelope disparity in our experiments were typically close to 2%. The luminance of the monitor was linearised by taking luminance measurements with a photometer, to which a logarithmic curve was fitted and a linear lookup table generated. The residual error from the fitted curve was no more than 0.2% of the luminance at any one of the sample points. The mean luminance was 37.7 cd $m^{-2}$.

Experiments were carried out in a darkened room. The only visible illumination originated from the monitor. A modified Wheatstone stereoscope was used to view the binocular images on a single monitor. The distance from the screen to the stereoscope was 44 cm. The stereoscope was adjusted so that the stereo image pairs were aligned using parallel viewing geometry. The visual extent of each monocular image was 7.9°. Image pixels were square with a width of 2.0 min. A fixation spot was used as a reference point to help keep vergence fixed.

### 2.3. Stimuli

The basic second-order stimuli were contrast-modulated sinusoidal gratings. Some examples are shown in Fig. 1A. The envelope, $0.5(1+mE(x,y))$, was an approximation to a square-wave grating formed by summing the first and third harmonics. Here, $m$ is the depth of modulation (between zero and one), and $E(x, y)$ is given by:

$$E(x,y)=\alpha\left[\sin(\omega x)+\frac{1}{3}\sin(3\omega x)\right] \quad (1)$$

where $\omega$ is the fundamental frequency of the envelope, and $\alpha$ is chosen so that $E$ oscillates between 1 and $-1$.

Given the envelope and a carrier, $C(x, y)$, the mathematical form of the left and right contrast-modulated test stimuli is:

$$T_l(x,y)=\frac{\mu}{2}\left[1+\frac{a}{2}(1+mE(x+d/2,y))C(x,y)\right], \quad (2)$$

$$T_r(x,y)=\frac{\mu}{2}\left[1+\frac{a}{2}(1+mE(x-d/2,y))C(x,y)\right]. \quad (3)$$

where $\mu$ is the mean luminance, $a$ denotes the contrast of the carrier prior to the contrast modulation, and $d$ is the positional disparity. The stimuli were visible only within a circular window as shown in Fig. 1. For

[1] The sidebands are the frequencies on either side of the carrier frequency at which there is nonzero power. These sidebands grow in amplitude with the depth of contrast modulation.

Experiment 1, the surrounds of the circular window were black as in Fig. 1. In Experiment 2, the surrounds were equal to the mean luminance. For all these experiments, $C(x, y)$ was a horizontal sinusoidal grating. Its spatial frequency was either 2.25 or 4.5 cycles per degree (cpd).

### 2.4. Procedure

Subjects were seated with their heads stabilized in a chin rest in front of the Wheatstone stereoscope. They responded using a computer mouse in forced-choice discrimination tasks. In Experiment 1 subjects were asked to respond as quickly as they could, but were not constrained by the viewing time otherwise. In Experiment 2, the adaptation experiment, the test image was presented for 500 ms following the adapting grating.

## 3. Experiment 1

The purpose of this experiment was to determine whether a pre-cortical nonlinearity can account for second-order stereopsis. Towards this end, we examined two models. The first model involves an early compressive nonlinearity like that given by the Naka–Rushton receptor equation (Scott-Samuel & Georgeson, 1995). As discussed in Appendix A (also see Burton, 1973; Henning et al., 1975; Brown 1995; Scott-Samuel & Georgeson, 1995), a compressive nonlinearity applied to a contrast-modulated signal introduces a distortion product at the frequency of the contrast envelope. The distortion product is 180° out of phase with the envelope, and its amplitude at the envelope frequency increases as a function of modulation depth and the square of image contrast.

An alternative model posits that the strength of the envelopes disparity signal depends upon the amplitude of the carriers sideband frequencies. The amplitude of the sidebands is a measure of the energy in the second-order signal. Eq. (11) of the appendix shows that this model predicts that the amplitude of the envelope signal increases linearly as a function of modulation depth and contrast. These predictions are different from the compressive nonlinearity model because of the different linear coefficients that govern the effect of modulation depth, and because of the linear versus quadratic dependence on contrast.

To decide between the two models, we exploited the depth asymmetry found for contrast envelopes (Fleet & Langley, 1994b; Hibbard et al., 1995; Hibbard, 1997; Langley et al., 1998). We measured the contrast of an additive contribution of $E(x, y)$ that was required to override the depth asymmetry, so that $E(x, y)$ could be seen transparently at a different depth behind, or in front of, the carrier. The stimuli were similar to those in Eqs. (2) and (3), but with an additive contribution of $E(x, y)$:

$$T_l(x, y) = \frac{\mu}{2}\Big[1 + bE(x + d/2, y) + \frac{a}{2}(1 + mE(x + d/2, y))C(x, y)\Big], \quad (4)$$

$$T_r(x, y) = \frac{\mu}{2}\Big[1 + bE(x - d/2, y) + \frac{a}{2}(1 + mE(x - d/2, y))C(x, y)\Big]. \quad (5)$$

where $\mu$ is the mean luminance, $b$ is the contrast of the additive contribution of $E(x, y)$, $a$ is the mean contrast of the original contrast-modulated term, and $a$ and $b$ are constrained so that $a + b < 1$. In different sessions, we varied the modulation depth $m$ of the envelope so we could test the predictions of each model. The range of $m$ that we tested is shown in Fig. 2B. In order to have a wide range of contrasts available at which we could test for the threshold of $b$, the contrast of the carrier, $a$, was fixed at 0.3. The spatial frequencies of $E(x, y)$ and $C(x, y)$ were 0.45 and 4.5 cpd, respectively. Note that while a distortion product introduces power that is 180° out of phase with the envelope, the additive signal here is in-phase with the envelope.

In each trial, subjects reported whether the combined envelope and luminance signal (referred to below as the hybrid envelope) was seen behind the horizontal grating. When $b$ was sufficiently large, subjects reported that the horizontal carrier was perceived in the depth plane of the fixation spot (Wurger & Landy, 1989; Langley et al., 1998), with the hybrid envelope appearing transparently behind it. On the other hand, when $b$ was small, subjects reported a flat (coherent) surface. This can verified by cross-eyed fusion of Fig. 1A. The disparity of the envelope was uncrossed and fixed at 20 min. At this disparity, with $b = 0$, all subjects reported a flat (coherent) surface.

Within a single session (for a fixed value of $m$), the contrast of the additive signal, $b$, was controlled from trial to trial in an adaptive fashion using APE (Watt & Andrews, 1981). Psychometric functions were fitted to the data. The magnitude of $b$ at which the subjects reported that the carrier lay in front of the hybrid envelope on 50% of the trials was taken as the contrast threshold. Each session was repeated three times to obtain threshold values for each condition.

## 4. Results and discussion

Fig. 2 summarizes the results. Fig. 2A shows JBs

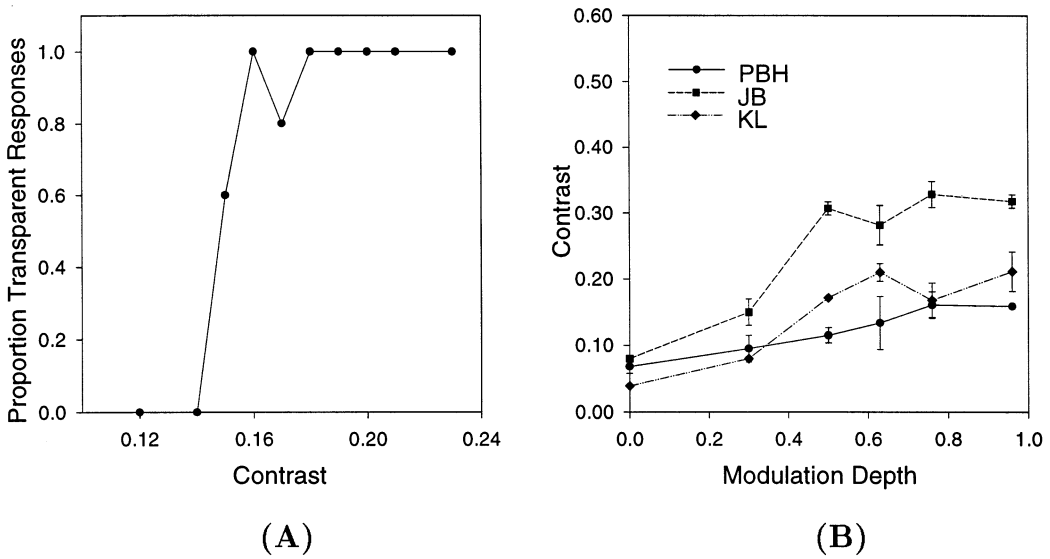


Fig. 2. (A) A psychometric function from subject JB shows the probability of perceiving transparency as a function of the contrast, $b$, of the added signal. The envelope modulation depth was 0.5. The threshold represents the value of $b$ required by subjects to see transparency on 50% of the trials such that the hybrid envelope was reported to lie behind the carrier. The psychometric function was generated using APE. It should be noted that during each session, the number of trials at each contrast level would have been different because APE is an adaptive method. (B) Transparency thresholds are shown as a function of modulation depth. The error bars represent the standard error of subjects 50% thresholds over three different sessions.

psychometric function for one session, when $m = 0.5$. As the contrast $b$ increased, the percept of transparency (i.e. ones ability to observe the vertical structure at a different depth than the horizontal structure) became more likely. Fig. 2B shows the contrast of the additive signal that was required to perceive the hybrid envelope behind the carrier on 50% of the trials, averaged over all three subjects. Notice that this contrast threshold increases with modulation depth. For the modulations depths studied, a linear regression, weighted by the inverse variance from each subjects measurements, yielded slopes of $0.23 \pm 0.17$ (KL), $0.29 \pm 0.21$ (JB), and $0.10 \pm 0.04$ (PB) with intercepts of $0.04 \pm 0.07$, $0.09 \pm 0.10$, and $0.07 \pm 0.02$, respectively[2]. The mean slope averaged across all three subjects, again weighted by the inverse variance of each subjects measurements, was found to be $0.211 \pm 0.09$.

We now consider which of the models discussed above might account for the data in Fig. 2. In particular, we consider one model based upon distortion products introduced by an early Naka–Rushton nonlinearity like that proposed by Scott-Samuel and Georgeson (1995). We also consider a model in which the contrast envelope is explicitly extracted by, for example, a quadratic nonlinearity often used in an energy mechanism. The appendix explains some specific predictions of these two models that are used below.

[2] Each measurement reported here is given with a 95% confidence interval taken from a two-tailed students $t$-distribution.

The first problem with the early nonlinearity model concerns the expected amplitude of the distortion product. We applied the Naka–Rushton nonlinearity to our stimuli with the parameters determined by Scott-Samuel and Georgeson (1995) and found that the amplitudes of the distortions products were too small. For example, in the condition shown in Fig. 2, with $m = 0.5$ and $a = 0.3$, the contrast of the envelopes distortion product was estimated at 0.007. However, a 0.45 cpd luminance grating alone at this contrast is insufficient to yield a depth percept. By comparison, Fig. 3B shows that the mean disparity thresholds when the modulation depth was zero (which is a simple additive transparency condition as shown in Fig. 1B) was $0.06 \pm 0.04$. Therefore, the amplitude of the distortion products introduced by the Naka–Rushton nonlinearity are below the contrast threshold required to produce a reliable disparity signal for the stimuli used in our experiments.

Eq. (10) in the appendix also shows that, according to this form of nonlinearity, the contrast $b$ required to override the asymmetry should increase as a function of modulation depth with a slope close to 0.0144. As calculated from Fig. 2, the mean slope for the three subjects was $0.21 \pm 0.09$. This value departs significantly the predicted value of 0.0144 that was obtained from the pre-cortical nonlinearity model.

Finally, concerning the pre-cortical nonlinearity model, it is interesting to note that the distortion

product and the additive signal $bE(x, y)$ are expected to be 180° out of phase. Therefore one might expect that there should be a value of $b$ for which they combine destructively and cancel. In this case one would expect no resultant disparity cue and a loss in depth sensation. We found no circumstances in which added luminance information nulled the depth signal in the contrast envelope. This was also the case when we added luminance information to only one eye's input. By similar arguments, an early compressive nonlinearity would not account for the data reported by Lin and Wilson (1995). They showed that a binocular depth signal could be detected between a luminance and contrast defined D6 Gaussian. These data may not be predicted from a pre-cortical nonlinearity model because the D6 Gaussian distortion products would be out-of-phase with the luminance defined signal. Unlike periodic signals, a disparity signal from two phase-reversed Gaussian functions may not be detected by a first-order process.

Instead of looking to a pre-cortical nonlinearity to explain the data in Fig. 2B, one might consider a cortical nonlinearity that extracts the modulating envelope after band-pass filtering. With such a model, one can show that the amplitude of the modulating envelope is a linear function of the amplitude of the carrier sidebands. As discussed in the appendix, one can show that the sideband amplitudes will vary linearly with the contrast $a$ and with the depth of modulation $m$. In particular, with $a = 0.3$ in the experiment here, it predicts that the sideband amplitudes should increase linearly as a function of $m$ with a slope of 0.138. This is much closer to the average slope of 0.21 found from Fig. 2B than the slope predicted by the early nonlinearity model above.

This view is also consistent with the predictions obtained from Metellis' (1974) monocular constraints on transparency; namely, that the product of two positive-valued functions may be perceived symmetrically in front of, or behind each other. With our stimuli it is easy to show that, when $b = am/2$ the stimulus in Eq. (3) is equivalent to a product of two positive-valued signals. In this case:

$$T_l(x, y) = \frac{\mu}{2}\left[1 + bE(x + d/2, y) + \frac{a}{2}(1 + mE(x + d/2, y))C(x, y)\right],$$
$$= \frac{\mu}{2}\left[\left(1 - \frac{a}{2}\right) + \frac{a}{2}(1 + mE(x + d/2, y))(1 + C(x, y))\right]. \quad (6)$$

When $b = am/2$ the signal in Eq. (6) is equal to a product of two-positive valued signals. Moreover, this relationship between $b$, the depth of modulation $m$, and the contrast $a$, is very close to that in Eq. (11) that describes the amplitude of the signals sideband frequency components (after the signal has been compressed by a Naka–Rushton nonlinearity). It provides another way to predict the threshold on $b$ at which transparent depth perception occurs (see also Langley & Hibbard, 1994).

Finally, it is worth noting that Langley et al. (1998) showed that transparent depth perception for additive combinations of the same signals used here are perceived symmetrically. In addition to those results, the experiments reported here show that the product of the same two signals (when each is defined as a positive-valued function) may also be perceived symmetrically transparent. These observations rule out the possibility that the depth asymmetry found for contrast envelopes occurs solely because of different additive versus multiplicative combination rules that were used to construct the stimuli. It points to our hypothesis that binocular asymmetries reflect a property of second-order processing, consistent with a two-channel model.

## 5. Experiment 2

In our second experiment, subjects were first adapted to a high contrast sinusoidal grating, and then asked to report the relative depth of the envelope in a contrast-modulated test stimulus. The premise behind the experiment was that, if the site of adaptation was after the nonlinearity, then the envelope signal would be present (as a first-order signal), and one would therefore expect that the most effective adapting frequencies would be close to the envelope frequency. If the site of the adaptation was before the nonlinearity, then the envelope signal remains implicit in the sideband frequencies near the carrier. Therefore we would expect that adaptation to frequencies near the carrier would affect the perceived depth of the envelope significantly. In addition to these two factors, if the adaptation was strongly orientation- and frequency-specific, then this implies that it occurs in visual cortex. In this case one might predict that contrast thresholds (as a function of the adapting gratings frequency) would show the same frequency and orientation dependent tuning as simple cells found in area V1 of the visual cortex (Hubel, 1988; De Valois & De Valois, 1990).

To explore these ideas, we used test stimuli that were the same as those defined in Eqs. (2) and (3). The depth of modulation $m$ was fixed at unity, while the contrast $a$ was varied as an independent variable. In these experiments, the spatial frequency of the carrier grating and the contrast envelope were varied in different sessions.

Subjects were adapted to a 98%-contrast sinusoidal grating for 2 min. The grating was counterphase flickered at 4 Hz to reduce phase-dependent after effects

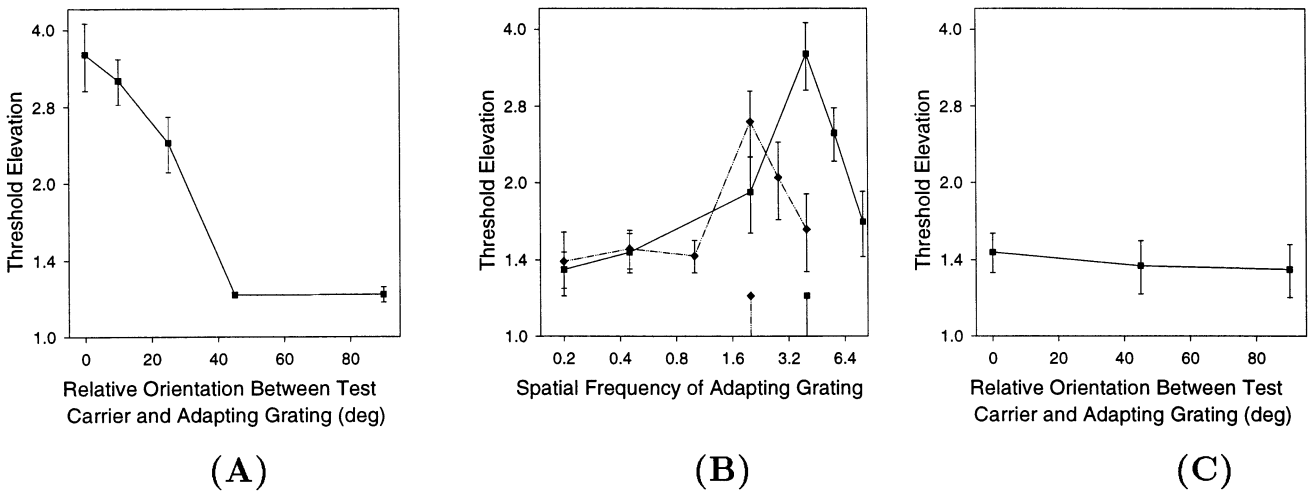


Fig. 3. (A) Threshold elevations are shown as a function of the angle between the carrier and adapting grating. The frequencies of the carrier and adapting gratings were identical. (B) Threshold elevations as a function of the adapting frequency. The carrier was parallel to the adapting grating. The two carrier frequencies are indicated on the horizontal axis. (C) Threshold elevations as a function of the angle between the carrier and adapting grating. The frequencies of the adapting grating and the envelope were equal. Error bars represent the standard error taken calculated from each subjects mean thresholds.

(Georgeson, 1987). Subjects were then shown a binocular test stimulus like that in Fig. 1A for 500 ms and asked to report whether the contrast envelope was in front or behind a fixation spot. This was immediately followed by a top-up adaptation period of 6 s, followed by another test stimulus on each subsequent trial. The contrast of the test stimulus ($a$ in Eqs. (2) and (3)) was varied using the method of constant stimuli.

Baseline contrast thresholds were measured before each session by repeating the same task without an adapting grating. These data were obtained so that we could compute threshold elevations after adaptation as a multiple of the baseline threshold.

A logistic function (ranging from 0.5 to 1) was fitted to the data taken from each subject in each session. From the resulting psychometric curves, 75%-correct disparity thresholds were measured. Each measurement was divided by the 75%-correct disparity thresholds taken from the baseline task to obtain the threshold elevations. Each session consisted of test images presented at ten different contrast levels, eight times each.

Each session was repeated three times, yielding 24 trials at each contrast level. Three sets of sessions were run:

- In the first set of sessions the frequency of the adapting grating was equal to the carrier frequency in the test stimulus, and we varied the angle between the adapting grating and the carrier grating. This allowed us to examine the importance of orientation when adapting frequencies were close to the carrier. We used angles of 0, 10, 25, 45, and 90°. The carrier frequency was 4.5 cpd. The fundamental frequency of the contrast envelope was 0.45 cpd.
- In the second set, the spatial frequency of the adapting grating was varied, but the orientations of the carrier and adapting grating were identical. Two carrier frequencies (2.25 and 4.5 cpd) were used to show spatial frequency tuning. The frequencies of the adapting gratings differed from the carrier by factors of, 0.5, 1.0, 1.414 and 2. We also used an adapting frequency that was equal to the fundamental frequency of the envelope and one that was 1 octave lower than the envelope frequency.
- In the third set, the frequency of the adapting grating was equal to the fundamental frequency of the envelope. Then we varied the angle between the adapting grating and the carrier, using angles of 0, 45 and 90°. Because the carrier and envelope are perpendicular, when the orientation between the adapting grating and carrier was 90°, the adapting grating and envelope were parallel.

## 6. Results and discussion

Fig. 3 shows the threshold elevations averaged across all duplicate sessions and all three subjects. Threshold elevations were greatest when the adapting and carrier signals were similar in frequency and orientation. These trends are apparent from Fig. 3A and B. Fig. 3C shows that threshold elevations remained small when we adapted to the fundamental frequency of the envelope rather than the frequency of the envelopes carrier grating. The figure suggests that threshold elevations were affected by the orientation of the adapting grating because the highest elevation occurred when the orien-

tation of the adapting grating was equal to the carrier. This trend in our data was, however, statistically insignificant.

If the adaptation occurred after the principal nonlinearity of second-order processing, then one would not expect threshold elevations to depend strongly on small differences between the adapting grating and the carrier. Instead, one might have expected the difference between the adapting grating and the envelope to have had a more a significant impact. A nonlinearity early in the visual system would introduce Fourier energy at the envelope frequencies, which could then be processed by luminance channels tuned to the envelope rather than the carrier. The data do not agree with these predictions; the curve shapes exhibit orientation- and frequency-selective tuning that depend upon the carrier frequency. The data suggest that the luminance signal was processed by orientation and frequency-selective channels before the contrast envelope was detected, and that the nonlinearity is therefore cortical. Also, note that because the carrier in the test stimulus was oriented horizontally these data imply that channels tuned to horizontal orientations play a role in binocular matching.

Threshold elevations decay from their maxima in Fig. 3A and B by a factor of two after 25° in orientation and 1 octave in frequency. Similar results have been reported in spatial vision, following contrast adaptation to sinusoidal gratings (De Valois & De Valois, 1990). These tuning curves are often attributed to orientation-selective neurons in the primary visual cortex (Hubel, 1988). These trends that further support the idea of a cortical nonlinearity.

Similar results were obtained from an analogous study on the effects of adaptation to spatial contrast envelopes that were modulated by sinusoidal gratings (Fleet & Langley, 1994b; Langley et al., 1996). These previous experiments concerned contrast thresholds for the detection of envelope orientation. Like the data reported here, those experiments found that adaptation to the envelopes carrier had the greatest influence on contrast thresholds for orientation discrimination of the envelope. One can note, however, that Langley et al. did not test for adaptation at the spatial frequencies of the contrast envelope, while these experiments did. There was, however, one notable difference between the data reported by Langley et al. (1996) and those found here. In the spatial adaptation paradigm, the threshold elevations for contrast envelope's reported by Langley et al. (1996) were almost an order of magnitude larger than the ones found here. There are two plausible explanations for this difference. One was that baseline thresholds for the envelope disparity task used here were higher than those for envelope orientation discrimination. Thus, in the experiments reported here there was less of a difference between the adapting contrast and the test contrast. Another explanation may be the disparity selectivity of adaptation, where adaptation to a zero disparity grating mainly affects the detection of disparities near zero (Blakemore & Hague, 1972). In the experiment here the adapting pattern was at zero disparity but the test patterns had crossed or uncrossed disparities. So one might again expect somewhat weaker effects following contrast adaptation to an envelopes carrier grating.

## 7. Conclusions

The data presented in this paper suggest that the site of the principal nonlinearity in second-order stereopsis is cortical. Experiment 1 showed that the amount of additive Fourier power at envelope frequencies required to over-ride the depth asymmetry reported by Langley et al. (1998) is inconsistent with a pre-cortical nonlinearity like that examined by Scott-Samuel and Georgeson (1995). Rather, it is consistent with an influence of Metelli's monocular constraints on binocular transparency, and with the energy contained in the sidebands of the contrast modulated stimuli. Experiment 2, showed that adaptation to the orientation and spatial frequency of the carrier has a more significant impact on one's sensitivity to the contrast envelope. These results suggest that the significant nonlinearity used to extract the envelope information followed orientation and frequency selective filtering and occurred in the visual cortex.

In light of this, one remaining question concerns the purpose of the early nonlinearities reported by other researchers (Burton, 1973; Henning et al., 1975; Scott-Samuel & Georgeson, 1995). One possible account is based on the anatomical constraints of transmitting retinal image signals via the LGN and optic nerve to the primary visual cortex. The transmission of information by neurons is limited by three factors: (i) transmission bandwidth, (ii) spike rate and (iii) shot noise (Laughlin, 1994). To transmit a faithful representation of the retinal image, Laughlin (1994) proposed that vertebrate visual systems optimise the coding of contrast at the earliest opportunity. One optimisation technique may be to introduce a nonlinear (say logarithmic or Naka–Rushton) compression prior to transmission. This process would offer a higher signal to (quantization) noise ratio in comparison to a linear quantization, especially when transmission bandwidth is limited (e.g. Schwartz, 1987). Following such compression, an inverse transformation is necessary to recover the original signal. Such inverse transformations are generally ill-conditioned and hence noisy. Inverse transformations of this form may also introduce expansive nonlinearities into the processed signal (Schwartz, 1987). Whether these inverse transformations occur is a potential topic of further research.

This perspective suggests that both expansive and compressive signal nonlinearities may be present in the visual cortex, but that their sources would be different. One would expect compressive nonlinearities to have been introduced early, soon after reception of the visual signal by the photoreceptors (cf. Burton, 1973; Sclar et al., 1990; Scott-Samuel & Georgeson, 1995). Conversely, expansive nonlinearities would be apparent much later, perhaps in the cortex. Harris and Smallman (1995) found evidence for expansive nonlinearities which they thought to be cortical in origin. These expansive nonlinearities could reflect imperfections that occurred during the encoding, transmission and reception of the signals from the retina.

Given that the significant nonlinearity used to detect second-order signals is cortical, there remain two models to consider. There are single-channel models that exploit an explicit logarithmic nonlinearity after orientation and frequency selective filtering (Langley, 1999). There are also two-channel models in which first and second-order signals are processed by independent first-order filters (Liu et al., 1992; Sato & Nishida, 1993; Hess & Wilcox, 1994; Sato & Nishida, 1994; Hibbard et al., 1995; Lin & Wilson, 1995; Wilcox & Hess, 1995, 1996, 1997). These two-channel models of binocular matching resemble those used to explain perception for second-order motion (Chubb & Sperling, 1988; Wilson et al., 1992; Zanker, 1993; Fleet & Langley, 1994a).

The model proposed by Langley (1999) involved a logarithmic transformation in order to linearise multiplicative combinations of image signals. Provided that the mean luminance of image has been removed by some pre-filtering, the logarithm transforms the nonlinear product of a contrast envelope and carrier signal into a sum of these two signals. Following a transformation of this form, it is feasible that a single model of binocular depth detection could then be used to explain binocular matching for both first and second-order signals as in the models proposed by Weinshall (1991) and Parker, Johnston, Mansfield and Yang (1991).

This model might predict that the visual system has the capacity to detect binocular signals that are corrupted by multiplicative noise (e.g. Franks, 1969). It might also explain some results of other researchers who have used similar stimuli to identify properties of the postulated second-order channel (e.g. Liu et al., 1992; Sato & Nishida, 1993; Lin & Wilson, 1995; Wilcox & Hess, 1995, 1996, 1997). However, this single-channel model would not explain depth asymmetries reported for second-order signals (Frisby & Mayhew, 1978; Halpern, 1991; Kersten, 1991; Sato & Nishida, 1994; Hibbard et al., 1995; Langley et al., 1998). This is because this single-channel model does not determine the origin of each disparity cue, whether first or second-order. Using the same reasoning, a pre-cortical nonlinearity model would also find it difficult to account for the depth asymmetry.

Taken together with the results of others, the results reported here support the hypothesis that, for stereopsis, contrast envelopes are detected by a cortical nonlinearity, and that first and second-order signal combinations may be processed by different mechanisms.

## Acknowledgements

This work was financially supported by grants to KL and PBH from the Graduate School and Department of Psychology at UCL, and by a grant from NSERC Canada and an Alfred P. Sloan Research Fellowship to DJF. We are grateful for comments and suggestions provided by two reviewers, and to H. Spekreijse and E. Borghols for their patience.

## Appendix A. Naka–Rushton receptor nonlinearity

Scott-Samuel and Georgeson (1995) have measured the magnitude of early visual nonlinearities (also see Brown (1995)). They temporally interleaved two spatial stimuli, one equal to the product of two gratings, and one equal to the sum of the same two gratings. In successive presentations of the product (every second frame), the contrast envelope was phase shifted by 180°. So there is no motion cue from the envelope alone. Similarly, in each successive presentation of the sum, the phase of the envelope component was phase shifted by 180° to remove motion cues in the sum signal. Finally, successive presentations of the envelope in the product were phase shifted 90° from the corresponding component in the sum.

An early compressive nonlinearity applied to the contrast-modulated signal produces a distortion product, containing energy at the envelope frequencies. Moreover, there are only 90° phase shifts between a distortion product in one frame and the corresponding Fourier components of the sum in the next frame. This gives rise to a first-order motion signal. In their experiments, Scott-Samuel and Georgeson varied the amplitude of the envelope components in the sum to find the contrast threshold at which this first-order motion signal could be nulled.

To model the early nonlinearity they used a Naka–Rushton receptor equation (e.g. Sclar et al., 1990) the response of which is given by:

$$R(x) = \frac{T(x)}{T(x) + S} \tag{7}$$

where $S$ is a semi-saturation constant, and $T(x)$ the input signal (scaled to have a maximum value of one). This model was found to give a good fit to their contrast threshold data with semi-saturation constants

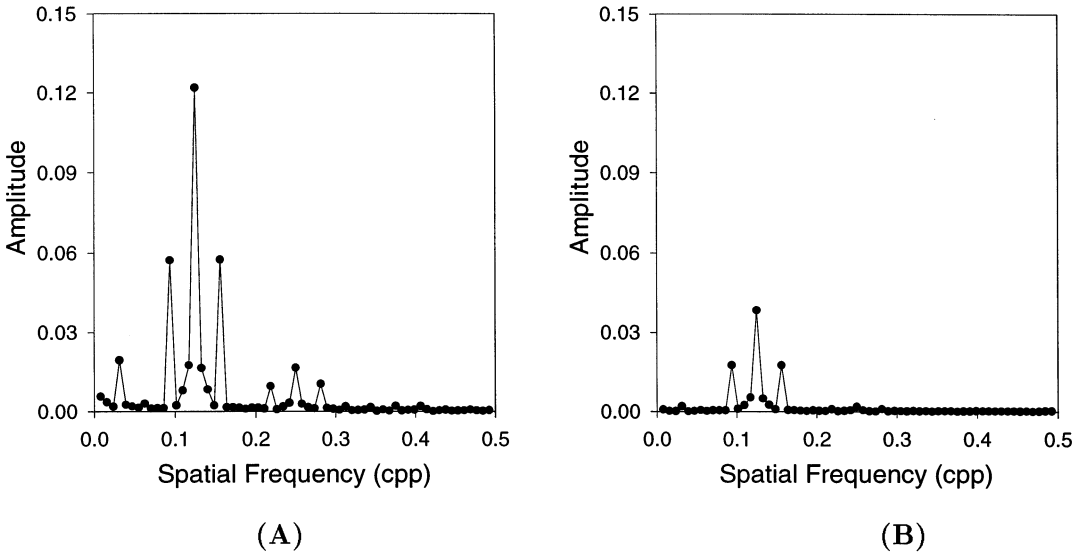


Fig. 4. Fourier spectrum (cycles per pixel) for the contrast modulated signal in Eq. (9) after being passed through a Naka–Rushton compressive nonlinearity. (A): Spectrum obtained for contrast-modulation depth and a semi-saturation constant of unity. (B) Spectrum obtained by reducing the mean contrast of the signal by one-third. In comparing the figures, note that the magnitude of the distortion product is reduced by changes in contrast as predicted by Eq. (11).

between 1.0 and 2.8. Scott-Samuel and Georgeson also found that the amplitude of distortion products increased with increasing temporal frequency. As noted by Sclar et al. (1990), the responses of motion sensitive neurons within magnocellular areas of the macaque LGN show greater compressive responses at high contrast than cells in parvocellular areas. If this occurs in human visual processing also, then we may regard these estimates of early nonlinearities as somewhat conservative because our stimuli were static. In what follows we consider this model with $S=1$ as it produces the largest distortion products into the transformed signal.

The basic stimuli used in the experiments here were one dimensional, of the form:

$$T(x)=\frac{\mu}{2}\left[1+\frac{a}{2}(1+mE(x))C(x)\right] \tag{8}$$

To normalize this signal to have values between 0 and 1, we divide by $\mu$ to obtain:

$$T(x)=\frac{1}{2}+\frac{a}{4}(1+mE(x))C(x) \tag{9}$$

where $0\le a\le 1$ is the amplitude of the carrier grating, $E(x)$ is an approximation to a square-wave grating, and $0\le m\le 1$ is the modulation depth. To understand the effect of the nonlinearity, we apply the compressive nonlinearity in Eq. (7) to the signal given by Eq. (9) to obtain $R(x)$. The Fourier spectrum of $R(x)$ then gives us the spectrum of the transformed signal which includes the distortion product at the envelope frequencies.Fig. 4 shows the Fourier spectra of $R(x)$ in two cases, namely, when $a=1$ and when $a=1/3$ (where $S=1$ and $m=1$). From the Fourier spectrum, at the frequencies of the envelope, we measured the amplitude of the distortion product as a function of contrast a (between 0.3 and 1.0) and modulation depth m (between 0 and 1.0). From these measurements we found that the amplitude $D_b$ of the distortion product at envelope frequencies increases approximately linearly with modulation depth, and approximately quadratically with the carrier amplitude. Our fit to these data produced the following relation:

$$D_b=0.16\,a^2\,m \tag{10}$$

This equation allows us to calculate the expected amplitude of early nonlinearities introduced as a function of carrier contrast and modulation depth. In particular, with a contrast of $a=0.3$, like that used in Experiment 1, we might expect contrast thresholds that depend linearity on $m$ with a slope of 0.0144.

Alternatively, one might suppose that the critical source of information is the amplitude of the carriers sidebands. Using the same approach as above, after applying the nonlinearity to Eq. (9), we measured the amplitude $D_s$ of the sidebands in $R(x)$, and found that they vary linearly with $m$ and $a$, that is:

$$D_s=0.46\,a\,m \tag{11}$$

Note that if we had ignored the nonlinearity, then it is easy to show that the sideband amplitudes should equal $\frac{am}{2}$; with the nonlinearity it is slightly different.

A comparison of Eqs. (10) and (11) shows that the amplitude of the sidebands increases with a greater slope (as a function of modulation depth) than the amplitude of the distortion product introduced by the early nonlinearity.

## References


Blake, R., & Wilson, H. (1991). Neural models of stereoscopic vision. *Trends in Neuroscience*, *14*, 445–452.

Beck, J., Prazdny, K., & Ivry, R. (1984). The perception of transparency with achromatic colors. *Perception and Psychophysics*, *5*, 407–422.

Beck, J. (1984). Perception of transparency in man and machine. Human and Machine Vision II, 1–12.

Blakemore, C., & Hague, B. (1972). Evidence for disparity detecting neurons in the human visual system. *Journal of Physiology*, *225*, 437–455.

Brown, R. O. (1995). Luminance nonlinearities and 2nd order stimuli. *Investigative Ophthalmology and Visual Science* (*Suppl.*), *36*, 51.

Burton, G. J. (1973). Evidence for non-linear response process in the visual system from measurements on the thresholds of spatial beat frequencies. *Vision Research*, *13*, 1211–1255.

Chubb, C., & Sperling, G. (1988). Drift-balanced random-stimuli: a general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A*, *5*, 1986–2007.

Cormack, L. K., Stevenson, S. B., & Schor, C. M. (1991). Interocular correlation, luminance contrast and cyclopean processing. *Vision Research*, *31*, 2195–2207.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1995). Neuronal mechanisms underlying stereopsis: how do simple cells in the visual cortex encode binocular disparity? *Perception*, *24*, 3–32.

Derrington, A. M., & Badcock, D. R. (1986). Detection of spatial beats: nonlinearity or contrast increment detection? *Vision Research*, *26*, 343–348.

De Valois, R. L., & De Valois, K. K. (1990). *Spatial vision*. Oxford: Oxford University Press.

Fleet, D. J, & Langley, K. (1994). Computational analysis of non-Fourier motion. *Vision Research*, *34*, 3057–3079.

Fleet, D. J., & Langley, K. (1994). Non-Fourier channels in motion and stereopsis. *Perception* (*Suppl.*), *23*, 83.

Fleet, D. J., Wagner, H., & Heeger, D. J. (1996). Neural encoding of binocular disparity: energy models, position-shifts and phase-shifts. *Vision Research*, *36*, 1839–1857.

Franks, L. (1969). *Signal analysis*. NJ: Prentice-Hall.

Frisby, J. P., & Mayhew, J. E. W. (1978). The relationship between apparent depth and disparity in rivalrous-texture stereograms. *Perception*, *7*, 661–678.

Georgeson, M. (1987). Temporal properties of spatial contrast vision. *Vision Research*, *27*, 765–780.

Halpern, D. L. (1991). Stereopsis from motion defined contours. *Vision Research*, *31*, 1611–1620.

Harris, J. M., & Smallman, H. S. (1995). Visual distortion products from an expansive nonlinearity. *Perception* (*Suppl.*), *24*, 126.

Henning, G. B., Hertz, G. B., & Broadbent, B. E. (1975). Some experiments bearing on the hypothesis that the visual system analyses patterns in independent bands of spatial frequency. *Vision Research*, *15*, 887–897.

Hess, R. F., & Wilcox, L. M. (1994). Linear and non-linear filtering in stereopsis. *Vision Research*, *34*, 2431–2438.

Hibbard, P. B., Langley, K., & Fleet, D. J. (1995). Transparent asymmetry in stereopsis. *Perception* (*Suppl.*), *24*, 35.

Hibbard, P.B. (1997) Linear and nonlinear mechanisms in the perception of stereoscopic slant and transparency. Ph.D Thesis, Department of Experimental Psychology, University College London.

Hubel, D. H. (1988). *Eye, brain and vision*. New York: Scientific American Books.

Kersten, D. (1991). Transparency and cooperative computation of scene attributes. In M. Landy, & J. A. Movshon Jr, *Computational models of visual processing* (pp. 209–228). London: MIT Press.

Langley, K., & Hibbard, P. B. (1994). Multiplicative/additive combination rules towards transparent stereoscopic slant perception. *Perception*, *23*, 33.

Langley, K. (1997) Degenerate models of multiplicative and additive motion transparency, vol. 2 (pp. 440–450). Proceedings of the British Machine Vision Conference, Colchester, UK.

Langley, K. (1999). Computational models of coherent and transparent plaid motion. *Vision Research*, *39*(1), 87–108.

Langley, K., Fleet, D. J., & Hibbard, P. B. (1996). Linear filtering precedes nonlinear processing in early vision. *Current Biology*, *6*, 891–896.

Langley, K., Fleet, D. J., & Hibbard, P. B. (1998). Linear and nonlinear transparencies in stereopsis. *Proceedings of the Royal Society of London B*, *265*, 1837–1845.

Laughlin, S. B. (1994). Matching coding, circuits, cells and molecules to signals: general principles of retinal design in the flys eye. *Progress in Retinal and Eye Research*, *13*, 165–195.

Liu, L., Schor, C. W., & Ramachandran, V. S. (1992). Positional disparity is more efficient in encoding depth than phase disparity. *Investigative Ophthalmology and Visual Science* (*Suppl.*), *33*, 1373.

Lin, L., & Wilson, H. R. (1995). Stereoscopic integration of Fourier and non-Fourier patterns. *Investigative Ophthalmology and Visual Science* (*Suppl.*), *36*, 364.

Mallot, H. A., Arndt, P. A., & Bulthoff, H. H. (1995). Matching versus correlation in human stereopsis. *Biological Cylbernetics*, *72*, 279–293.

Mareschal, I., & Baker, C. L. Jr (1998). A cortical locus for the processing of contrast defined contours. *Nature Neuroscience*, *1*, 150–154.

Metelli, F. (1974). The perception of transparency. *Scientific American*, *230*, 90–98.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science*, *249*, 1037–1040.

Poggio, G. F., & Poggio, T. (1984). The analysis of stereopsis. *Annual Review of Neuroscience*, *7*, 379–412.

Parker, A. J., Johnston, E. B., Mansfield, J. S, & Yang, Y. (1991). Stereo, surfaces, and shape. In M. S. Landy, & J. A. Movshon, *Computational models of visual processing* (pp. 359–381). London: MIT.

Sato, T., & Nishida, S. (1993). Second-order depth perception with texture-defined random-dot stereograms. *Investigative Ophthalmology and Visual Science* (*Suppl.*), *34*, 1438.

Sato, T., & Nishida, S. (1994). Does an envelope detecting mechanism mediate stereopsis for band-limited stimuli? *Investigative Ophthalmology and Visual Science* (*Suppl.*), *35*, 1916.

Sclar, G., Maunsell, J. H. R., & Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research*, *30*, 1–10.

Scott-Samuel, N. E., & Georgeson, M. (1995). Does an early nonlinearity account forsecond-order motion? *Perception* (*Suppl.*), *24*, 104.

Schwartz, M. (1987). Information transmission, modulation and noise (3rd ed). London: McGraw-Hill.

Turano, K., & Pantle, A. (1989). On the mechanism that encodes the movement of contrast information. *Vision Research*, *29*, 207–221.

Victor, J. D., & Conte, M. M. (1992). Coherence and transparency of moving plaids composed of Fourier and non-Fourier gratings. *Perception and Psychophysics*, *52*, 403–414.

Watt, R., & Andrews, D. P. (1981). Adaptive probit estimation of psychometric functions. *Current Psychological Reviews*, *1*, 205–214.

Weinshall, D. (1991). Seeing 'ghost' planes in stereo vision. *Vision Research*, *31*, 1731–1749.

Wilcox, L. M., & Hess, R. F. (1995). $D_{max}$ for stereopsis depends on size, not spatial frequency content. *Vision Research*, *35*, 1061–1069.

Wilcox, L. M., & Hess, R. F. (1996). Is the site of nonlinear filtering in stereopsis before or after binocular combination. *Vision Research*, *36*, 391–399.

Wilcox, L. M., & Hess, R. F. (1997). Scale selection for second-order (non-linear) stereopsis. *Vision Research*, *37*, 2981–2992.

Wilson, H. R., Ferrera, V. P., & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, *9*, 79–97.

Wurger, S. M., & Landy, M. S. (1989). Depth interpolation with sparse disparity cues. *Perception*, *18*, 39–54.

Zanker, M. J. (1993). Theta motion: a paradoxical stimulus to explore higher-order motion extraction. *Vision Research*, *33*, 553–569.