Disparity-tuned channels of the human visual system

LAURENCE K. CORMACK, SCOTT B. STEVENSON, AND CLIFTON M. SCHOR

1Department of Psychology, University of Texas at Austin, Austin
2School of Optometry, University of California, Berkeley

(Received June 15, 1992; Accepted October 29, 1992)

Abstract

Traditionally, it has been thought that the processing of binocular disparity for the perception of stereoscopic depth is accomplished via three types of disparity-selective channels—"near," "far," and "tuned." More recent evidence challenges this notion. We have derived disparity-tuning functions psychophysically using a subthreshold summation (i.e. low-level masking) technique. We measured correlation-detection thresholds for dynamic random-element stereograms containing either one or two surfaces in depth. The resulting disparity-tuning functions show an opponent-type profile, indicating the presence of inhibition between disparity-tuned units in the visual system. Moreover, there is clear inhibition between disparities of the same sign, obviating a strict adherence to near–far opponency. These results compare favorably with tuning functions derived psychophysically using an adaptation technique, and with the tuning profiles from published single-unit recordings. Our results suggest a continuum of overlapping disparity-tuned channels, which is consistent with recent physiological evidence as well as models based on other psychophysical data.

Keywords: Stereopsis, Disparity, Channels

Introduction

In 1801, Thomas Young suggested in his lecture to the Royal Society of London that there are three principle resonators—centered on the spectral red, yellow, and blue—each of which was "capable of being put into motion less or more forcibly by undulations differing less or more from a perfect unison . . ." With these words, Young not only provided the suspected trichromacy of vision with a theoretical framework which might today be termed "neurally cheap and computationally efficient" but he also introduced the concept of sensory channeling.*

The channeling of information along a specific stimulus dimension is now a ubiquitous concept in the study of sensory systems. Stimulus dimensions such as temporal frequency in audition (Green, 1979) and both luminance- and color-contrast spatial frequency in vision (Campbell & Robson, 1968; Switkes et al., 1988; see DeValois & DeValois, 1988, for a review) have been shown to be processed through parallel sensory channels by the early levels of the nervous system. Both of these dimensions, however, seem to be processed by a fairly large number of channels whose peak sensitivities form a continuum along the dimension in question. This stands in contrast to current models of chromatic processing which, as suggested by Young, are conventionally held to consist of three channels.† There is also evidence that motion-in-depth is processed by only four distinct channels (Regan & Beverley, 1978).

The earliest channel-based hypothesis of disparity processing invoked three "poops" of disparity detectors (Richards, 1970, 1971, 1972). This evidence was based upon data showing that large numbers of otherwise normal observers could be insensitive to either near (crossed) or far (uncrossed) disparities. Additional observers seemed to correctly encode the magnitude of retinal disparity, but were unable to determine its sign. Richards hypothesized that these selective insensitivities were due to the absence of one of the three pools ("near," "far," or "tuned") of disparity detectors.

Apparent neurophysiological confirmation of this hypothesis came from Poggio and Fischer (1977), who recorded from neurons in V1 and V2 of macaque cortex and found neurons with disparity-tuning properties very similar to those of the pools hypothesized by Richards (1971, 1972).

More recent studies have challenged a strict adherence to a trichotomy of tuning in the disparity domain, however. Poggio et al. (1988) recorded from many neurons in macaque cortex that defied categorization under the former near–far–tuned scheme, and were thus termed "tuned–near" or "tuned–far."

Reprints requests to: Lawrence K. Cormack, Department of Psychology, University of Texas at Austin, Meezes 330, Austin, TX 78712, USA.

*It is of interest to note that Young's talk preceded by some 40 years Muller's treatise on the Doctrine of Specific Nerve Energies, which, in hindsight, is probably the most obvious formulation of sensory channeling.

†At the receptor level, these correspond to the three (but see Neitz & Jacobs, 1990) cone types. Most probably at the bipolar layer and certainly by the ganglion cell layer, these have been transformed into "red", "green", "blue", and "luminance" channels. Nevertheless, the trichromacy or "three-channelness" is preserved.
Similarly, LeVay and Voight (1989), recording from neurons in the cat visual cortex, found no evidence for discrete categories of disparity tuning. Rather, they describe disparity tuning in cat visual cortex as occurring along a continuum.

Early computational studies of stereo vision have tended to focus on the "matching problem" in such a way that the disparity resolution is not dictated by cyclopean pools or channels, but rather by the sampling grain of either the "front end" of the model or of the cortical "Keplerian grid." The former corresponds to the spacing of local elements or intensity samples in models such as that of Prazdny (1985), Poliard et al. (1985) or the scale of the spatial filters in Marr and Poggio (1979). The latter corresponds to the spacing, in the disparity domain, of the cells of the Keplerian grid in models such as that of Nelson (1977). Other computational models (Nishihara, 1988; Cormack et al., 1991) have assumed that stereo localization is limited by some property (e.g. the value of the second derivative) of a continuous cross-correlation function and, as such, have assumed no explicit disparity-tuned channels. In modeling the psychophysical disparity increment data of Badcock and Schor (1985), however, Lehky and Sejnowski (1990) concluded that a continuous family of disparity-tuned channels was a necessary component of the model. Attempts using a small (e.g. three) number of channels could not produce data similar to those of Badcock and Schor (1985).

In general, psychophysical studies of channeling have employed one of three techniques (masking, adaptation, and subthreshold sumation) to derive "tuning functions," the form of which can then be related to properties of the underlying channels. One of these, the masking paradigm, is probably best known for its use in defining the critical bands in audition, although it has also been widely used in vision. The method of adaptation has been successfully used to demonstrate the presence of spatial-frequency channels (Blakemore & Campbell, 1969), and to derive spatial-frequency tuning functions for both luminance (DeValois, 1977) and chromatic (Switkes et al., 1988) gratings. It has also been widely used to explore receptor spectral sensitivities (Stiles, 1949) and color-opponent mechanisms (Krauskopf et al., 1982). The third method of subthreshold summation, has been used in the spatial-frequency domain (Kulikowski & King-Smith, 1973), but the interpretation of the results is somewhat difficult due to the fact that it yields tuning functions that are much narrower than those derived from adaptation studies. Subthreshold summation has been used successfully in audition (Scharf, 1970) and color vision (Thornton & Pugh, 1983; Kranda & King-Smith, 1979; Sperling & Harwerth, 1971). It was also the technique used in the classic study of Levinson and Sekuler (1975), who demonstrated the presence of independent channels in the human visual system tuned for direction of motion.

The logic of the subthreshold summation approach is illustrated in Fig. 1. In the top panel, a single stimulus is presented to a single channel. This could either be due to the location of the stimulus or to the fact that only the channel shown by the solid line is present. In any case, threshold will be reached when the strength of the stimulus multiplied by the sensitivity of the system at that disparity reaches some criterion amount (i.e. \( d' = 1.0 \)).

In the middle panel, a second stimulus is now added. If both stimuli act upon a single channel, then we expect threshold level performance when the sum of the products of the stimulus strengths and the sensitivity of the system at the stimulus disparities reaches the same criterion amount. In other words, if the two stimuli are acting upon the same channel, we expect complete summation of the stimulus strengths. Thus, the strength of either stimulus at threshold in the two stimulus condition will be much lower than that of a single stimulus presented alone.

In the lower panel, a second stimulus is also added. In this case, however, the two stimuli act primarily upon separate, independent channels. We now expect a reduction in threshold relative to the single-stimulus condition only by an amount corresponding to probability summation.

An additional possibility, which is that there exist multiple channels which are not independent, is not illustrated in the
Disparity-tuned channels

figure. Yet evidence for mutually inhibitory sensory channels is known to exist in both spatial vision (DeValois, 1977) and color vision (Thornton & Pugh, 1983; Kranda & King-Smith, 1979).

In this paper, we report tuning functions in the disparity domain derived with a subthreshold summation technique. With this technique, we first measure correlation thresholds (Cormack et al., 1991; Tyler & Julesz, 1978) for a single dynamic random-element stimulus presented alone (referred to as “baseline thresholds”). A graph of these thresholds as a function of disparity represents the envelope over which correlation detection occurs and, as such, is analogous to a contrast-sensitivity function in spatial vision. Correlation in the cyclopean domain is very analogous to contrast in the spatial domain. They are both intensive variables. In fact, correlation and contrast trade off in a quite predictable manner in determining the strength of a cyclopean stimulus (Cormack et al., 1991). Disparity, however, is an extensive dimension, analogous to spatial frequency in the spatial domain or to wavelength in the spectral domain. Thus, a stimulus with a disparity of 30 arcmin is not a “stronger” stimulus than one with a disparity of 0 arcmin any more than a stimulus of 700 nm is a “stronger” stimulus to the photopic visual system than one of 555 nm. In fact, in both cases, the former stimuli are the “weaker” (all else being equal) due to the fact that they are under a less-sensitive portion of the envelope of sensitivity.

Once we had measured baseline correlation thresholds, tuning functions were derived by comparing the baseline thresholds with thresholds obtained for the same stimuli in the presence of other dynamic random dot stimuli also presented at or near threshold. We show that these tuning functions agree quite well with those derived from a psychophysical experiment using an adaptation technique as well as published tuning functions from the single-unit work of Poggio et al. (1988).

Methods

Stimuli

The stimuli were dynamic random-element stereograms of 50% density. Unlike conventional stereograms, there was never a disparity-defined “form in depth” present. When the disparity was changed, the apparent depth of an entire fronto-parallel surface was changed.

A pseudo-random noise signal was hardware generated via shift registers running at 7 MHz and was displayed on a pair of matched TSD monitors (p4 phosphor, 60 Hz non-interlaced) viewed through a mirror haploscope. The viewing distance was 57.3 cm and the haploscope mirrors were adjusted for each subject to the corresponding convergence angle, thus obviating any mismatch between convergence and accommodation or “higher level” distance cues. Mean luminance was roughly 30 cd/m². The displays were viewed through 7-deg circular apertures in an otherwise black surround.

Horizontal disparities were produced by delaying the horizontal video sync to one monitor. This was accomplished via a programmable delay chip (Digital Delay Devices model PDU-13256-0.5), which allowed us to delay the noise stimulus to one eye in 0.5 ns (corresponding to 2 arcsec) increments.

Intercocular correlation (IOC) was simply the proportion of the dots which were “forced” to match (i.e., have the same luminance) in the two images; the remainder of the dots then had a 50% chance of matching. Thus, in a display which had an interocular correlation of zero, half of the dots in the right image were matched by dots in the left image. For an interocular correlation of +1, of course, the two images were identical.

The proportion of matching dots, then, is simply \( (IOC + 1)/2 \).

Intercocular correlation was varied through the use of two independent noise generators. To produce an interocular correlation of +1, the output of a single noise generator was sent to both monitors. To produce an interocular correlation of zero, each monitor was driven by a separate noise generator. Intermediate interocular correlations were produced by switching between the above two conditions at a very high rate (156.25 kHz) to allow the spatio-temporal integration of the visual system to render the stimulus identical with one in which interocular correlation is statistically determined on a dot-by-dot basis. This switching pulse was not synchronized to the video signal, nor was the switching rate sufficiently close to an even multiple of the frame rate, lest stationary or drifting bands of correlation be visible. Using this method, the interocular correlation is simply the duty cycle of the (rectangular wave) switching pulse.

Thus, to create an interocular correlation of 0.75, the duty cycle of the rectangular wave switching pulse was such that the subject was viewing a fully correlated display 75% of the space/time, and viewing an uncorrelated display for the remainder. Given the very high switching pulse frequency and the 60-Hz non-interlaced frame rate, however, the perception was one of a continuously present intermediate correlation; no perception of the correlation switching was present during the experiments.

For the two-plane condition, each surface was presented on alternate frames. Thus, if the stimulus for a particular run consisted of a surface at zero disparity and a surface at 10’ (arcmin) crossed disparity, the zero-disparity surface would be present on the odd frames (say) and the disparate surface would be present on the even frames of the 12-frame stimulus interval.

To yield thresholds which were directly comparable to those in the two-plane condition, baseline thresholds were run with the test surface interleaved with frames of zero interocular correlation. Thus, the test surface would be present on only the odd frames (say) of the 12-frame stimulus interval. In effect, this procedure halved the correlation of the test surface, but allowed us to numerically compare the thresholds obtained in the two-plane and baseline conditions.

A few words on the phenomenology of these stimuli are perhaps worth mentioning here. When an observer views a stimulus in which there is no interocular correlation at any disparity, his or her impression is almost one of looking into a fog or a dense swarm of insects; there is definitely volume but no real form or surface. When the interocular correlation of a single surface is then increased from zero to 1.0, the observer's perception is of a surface emerging within the fog until, finally, at very high (approaching 1.0) correlations, a very robust surface is seen and no hint of the surrounding volume (or fog) is seen. This is despite the fact that only the correlation at a single disparity has been changed, nothing about the stimulus at the rest of the disparities where the fog was perceived has been altered. This might in itself indicate that some sort of disparity domain inhibition is taking place (although see Tyler’s (1977) arguments concerning the symmetry of false targets in dynamic random element displays).

The percept of a two-plane stimulus is rather similar, and is illustrated in Fig. 2. When the interocular correlation is raised

†Two control experiments, reported in Cormack et al. (1991) further insured the validity of this method of producing variable IOC.
from zero, through the positive values, to 1.0 (Fig. 2; top and bottom, respectively), the observers percept is of two distinct surfaces emerging from within the fog. It should be noted that at our 60-Hz non-interlaced frame rate, no flicker or motion-in-depth is ever seen with the two-plane stimulus. In fact, we are able to cleanly display up to four planes "simultaneously" (Stevenson et al., 1989).

Procedure

Two of the authors served as subjects. They had normal or corrected to normal acuity, normal contrast sensitivity, and excellent stereopsis.

A temporal 2AFC-method (alternate forced-choice) of constant stimuli paradigm was employed. A block of trials consisted of one trial at each of 4–6 correlation levels, chosen to bracket the expected threshold based on pilot data. The order of trials was randomized within blocks. A run was composed of 30 blocks of trials at a single-stimulus configuration. A threshold correlation for a particular stimulus configuration was defined as 75% correct on a psychometric (Weibull) function fit to the data from five such runs.

![Transparent Stereoscopic Surfaces](image)

**Fig. 2.** Illustration of the phenomenology of the two-surface displays used in the experiments. When a low, suprathreshold correlation is present at two disparities simultaneously, observers perceive two "lazy" depth planes embedded within a cloud of visual noise. The visual noise is the perceptual manifestation of spurious or "ghost" matches. As the interocular correlation is raised, the surfaces become more salient and "solid," even though the distal surface is clearly visible through the proximal surface. The visual noise disappears from perceptual awareness, even though the same number of viable spurious matches is present.

A trial consisted of two stimulus intervals, roughly 200 ms (exactly 12 video frames) in duration, which were delimited by audible tones. The dynamic noise was continuously present but, during one of the two intervals, switched from zero interocular correlation at all disparities to some positive interocular correlation at one baseline condition or two (two-plane condition) disparities for 200 ms. The subject's task was to signal, by means of a key press, which interval contained some nonzero interocular correlation. The subject's response was followed by auditory feedback and initiated the next trial.

The subjects fixated a 12-arcmin wide "+" located at the center of the display which, to assure accurate convergence, was flanked above and below by 48 × 5 arcmin Nonius lines. All stimulus disparity values therefore reflect disparities relative to the plane of fixation, as defined by the Nonius horopter. Psychophysical experiments show that our observers are sensitive to Nonius offsets corresponding to between 15 and 30 arcsec of disparity in our stimulus (when fully correlated), and eye movement recording experiments (using an SRI generation V DVI eye tracker) show that vergence remains constant to within ±2 arcmin (standard deviation of the mean) during a 12-s recording period when viewing these stimuli (Stevenson et al., in preparation).

Although we had already obtained extensive baseline data on our subjects, and thus knew with a fair degree of certainty what the baseline thresholds would be at any given disparity, we re-ran all of our baseline thresholds in this study, interleaving them with the two-plane runs. We did this for two reasons. First, we wished to obviate any possible effects of threshold variation over time.§ And second, we needed to run baselines in which every other frame was uncorrelated (as described above in the subsection Stimuli). This allowed the comparison with the "two-plane" stimulus in which both planes were in fact at the same disparity. We predicted that this would result in baseline thresholds which were elevated by a factor of two compared to those previously obtained and this was, in fact, the case.

Baseline thresholds were run at disparities spanning the range from −30 arcmin (crossed or "near" disparity) to +30 arcmin (uncrossed or "far" disparity) in 5-arcmin increments. Two-plane stimuli were composed of a plane at one of five disparities (−20, −10, 0, +10, and +20 arcmin) and a second plane at one of the disparities used to measure baseline thresholds. A particular disparity combination will hereafter be referred to as a "condition." For example, in order to measure the interaction of a plane at 0 arcmin (i.e. a stimulus coincident with the plane of fixation) and a plane at −10 arcmin, a total of 15 thresholds would be measured; five thresholds at each baseline condition (0 arcmin alone and −10 arcmin alone) and five thresholds at the [0, −10] condition (in which a surface was present at both 0 and −10 arcmin). To fully characterize the interaction between a plane at 0 arcmin and planes and other disparities (i.e. to derive a tuning function for 0 arcmin), baseline conditions at all of the relevant disparities were run, as well as nine two-plane conditions in which one of the planes was at 0 arcmin: [0, ±30], [0, ±20], [0, ±10], [0, ±5], and [0, 0]. Thus, between 400 and

---

§For practiced subjects, there seems to be no significant variation of threshold over time. We have looked at the zero disparity baseline thresholds for two subjects (LKC and SBS) across a period of over a year, and found no variation other than that which is typical on a day-to-day basis. This amount of variation (roughly an octave) is also displayed by ideal observers which we have developed for this task, and thus reflects the statistical nature of our stimuli more than anything else.
500 runs (60,000 to 75,000 trials) composed a complete data set (five tuning functions) from a subject. Complete data sets were collected on two of the three subjects (LKC and SBS) while data for one tuning function were collected for the third. The data from SBS and LKC will be analyzed in detail.

In the two-plane conditions, the correlations of the planes at each of the two disparities were always covariated in equal multiples of their respective baseline thresholds. Thus, when the two-plane condition \(0, -10\) was being run, the plane at zero disparity was always at the same number of baseline threshold multiples as the plane at \(-10\) arcmin. Since thresholds on the horopter were invariably lower than those 10 arcmin off the horopter, the physical correlation of the \(-10\) arcmin plane was always higher than that of the plane at 0 arcmin.

**Data analysis**

Within a condition, the percent correct at each stimulus level was averaged across runs. This procedure yielded average psychometric data for each subject which were then fit with a Weibull function by minimizing \(x^2\). First, the psychometric data comprising the mean percent correct at each stimulus level was fit. Then, the psychometric data comprising the mean percent correct plus one standard deviation at each stimulus level was fit. The difference between the threshold estimate (defined as the 75% correct level) of the latter and the former was taken as an estimate of the variability of the threshold estimate. This procedure probably provided an overestimation of variability due to the fact that the deviation of each data point on the psychometric function was taken in the same direction, thus representing a “worst case scenario.”

To derive disparity-tuning functions, the data were first plotted in a coordinate system in which the \(y\) axis represents the correlation at the disparity of one of the two planes, and the \(x\) axis represents the correlation at the disparity of the other plane (cf. Kranda & King-Smith, 1979). Both axes are normalized to the baseline threshold for a single plane at the disparity of the axis.

In Fig. 3, an example is given in which the data from one subject from the \([-10, +10]\) condition are plotted in this coordinate system. Notice that the baseline thresholds for the planes at \(-10\) and \(+10\) arcmin plot at unity on the \(x\) and \(y\) axis, respectively.

The advantage of this coordinate system is that it is fairly easy to distinguish whether or not the stimuli are being processed by a single channel, multiple independent channels, or mutually antagonistic channels. If the stimuli are being processed by a single channel (cf. upper panel of Fig. 1), then channel activity will sum completely and data are expected to fall along the diagonal which has a slope of \(-1\) and intersects both axes at 1.0. If, however, the stimuli are acting upon two independent channels, then the only threshold reduction which one expects is that which is due to probability summation. Data which result from probability summation among independent channels are predicted to fall along a curvilinear contour between the diagonal of slope \(-1\) and the bounds of the unit square. The exact form of the predicted contour depends on the slopes of the psychometric functions (the \(\beta\) parameter from the Weibull fit) for the baseline measurements. Shown in this figure is the contour expected if \(\beta\) were 2.2 for both baseline threshold measurements. In general, the slopes of our psychometric functions were almost always between 1.8 and 2.2, so, for analytical purposes, we used the unit circle as our prediction for data which are due to probability summation between independent channels. Relevant issues concerning probability summation are discussed more fully in the Appendix.

For the purposes of obtaining tuning functions, we defined the “summation index” for a condition (disparity combination) to be the distance between the data point from the condition and the unit circle. In other words, the summation index was the radius of the unit circle minus the distance of the data point from the origin. Since the two planes were always presented at

---

**The unit square represents the prediction of thresholds based on detection by separate channels in which the noise is fully correlated, or the limiting case in which the psychometric functions are infinitely steep (i.e., no noise in either channel). In our experiments, the performance limiting noise at high contrasts is the noise in our stimuli (Coraccia et al., 1991) and is uncorrelated across disparities (by the nature of dynamic random-dot stereograms). Thus, the unit circle, and not the unit square, represents the prediction for detection thresholds if the two planes of the two-plane stimuli are acting upon independent channels.**
equal threshold multiples, the polar coordinates of the data points were always roughly equal to \((r, 45)\) and the summation index was simply equal to \((1 - r)\).

Tuning functions were then derived by plotting, for a plane at a given disparity, the summation index as a function of the disparity of the plane with which it was paired in the various conditions. The derivation of a tuning function is illustrated in Fig. 4. This tuning function represents, in a single graph, all of the conditions in which one of the two surfaces was at \(-10\) arcmin. The disparity of the second plane is given by the abscissa. The insets show three examples of the plots from which the tuning functions are derived. Consider the lower right inset (a replot of Fig. 3), which plots the threshold level correlation of a plane at \(+10\) arcmin against that of a plane at \(-10\) arcmin. For the purposes of the present study, one need only attend to the filled data point closest to the 45-deg diagonal\(^\dagger\) (the reason that the point is not exactly on the 45-deg diagonal simply reflects the fact that our initial estimates of the baseline correlation thresholds were slightly off). The distance from this data point to the curved line (which represents the prediction given independent channels) is taken as the summation index. This summation index is plotted as the +10-arcmin point on the tuning function. In this example, the data point is far outside any reasonable prediction of independence indicating substantial inhibition, thus the sign of the summation index is negative.

The solid curve fit to the tuning function data is a best-fit Gabor function, which is a sinusoid multiplied by a Gaussian window. There was no theoretical motivation for selecting this function in particular, only the bias to believe that behavioral functions change smoothly.

Results

Baseline correlation-sensitivity functions for subjects LKC and SBS are plotted in Fig. 5. These functions are extremely stable over time; the anomalous "bump" at roughly +20 arcmin for LKC was also present in baseline data collected for the same subject over a year prior to the collection of the present data set. These functions represent the envelope over which correlation detection occurs and, as such, are analogous to contrast-sensitivity functions in spatial vision. As in spatial vision, the challenge is to discover the nature of the channels (and their interactions) which make up this envelope of sensitivity.

As can be seen from these data, the maximum sensitivity corresponds to about 8% correlation. This means the subjects could
correctly discriminate (75% correct performance) an interval in which 54% of the dots matched from one in which only 50% of the dots matched. On either side of the horopter, log correlation sensitivity decreases as a roughly linear function of disparity. The maximum disparity at which the subjects can perform the task (i.e. the point at which sensitivity is equal to 1.0) was found in a previous study to be approximately $\pm 60$ arcmin (Stevenson et al., 1992).

Figs. 6 and 7 plot all of the derived tuning functions, fit with Gabor functions from subjects LKC and SBS, respectively. The arrows on the abscissa indicate the constant disparity of one of the two planes. The disparity of the second surface is given by the data points themselves.

The main feature that is evident in all of the tuning functions is the manner in which the degree and the sign of signal summation varies as a function of the separation of the two planes in depth. When the two planes share the same or similar disparities, maximum summation occurs. As the separation of the two planes is increased by a modest amount, summation decreases to that predicted by probability summation (which is equivalent to a summation index of zero). As the separation of the planes is further increased, clear inhibitory behavior is observed, as indicated by summation indices less than zero. This inhibitory interaction is not minor; in every case, except the zero-disparity tuning function for LKC, it is as great in magnitude as the summation which occurs when both planes share a common disparity.

In the case of the off-horopteral tuning functions, the most pronounced inhibition observed was in the direction of the horopter. This could, however, be due to one of two factors. The first, and less appealing, is that this reflects the fact that, due to the technical limitations involved in presenting multiple depth planes, we were unable to present correlations higher than 0.5 in any single plane. Thus, we were unable to test disparities reliably beyond 30 arcmin in either direction or, for subject SBS, beyond 25 arcmin in the uncrossed direction.

The second possibility, which is suggested in the data, is that the off-horopteral tuning functions are inherently odd symmetric in nature. This would indicate an underlying channel structure in which excitatory interactions are local about the disparity in question, with inhibitory interactions occurring more strongly between channels nearer to, or on the other side of, the horopter.

In the case of the $\pm 10$-arcmin tuning functions, the crossover from excitation to inhibition occurred at or near the horopter. In the case of the $\pm 20$-arcmin tuning functions, however, clear inhibition was observed on the same side of the horopter in every case.

Shown in Fig. 8 is the family of five tuning functions derived for subject LKC. These tuning functions, which are not fit with any function, illustrate nicely the points outlined above. It should be stressed that curves shown in this figure are not
the "channels" mediating stereopsis. They are tuning functions; psychophysical constructs which reflect properties of the underlying channel structure.

The zero-disparity tuning function is symmetric, with inhibition occurring at the largest disparities which we could test. The $\pm 10$-arcmin tuning functions have roughly the same bandwidth as the zero-disparity tuning function, yet they have become markedly odd-symmetric, with a much steeper dropoff occurring in the direction of the horopter. In the case of the $-10$-arcmin tuning function, it is particularly clear that the crossover from excitation into inhibition occurs before the horopter is crossed.

In the $\pm 20$-arcmin tuning functions, we seem to see an increase in the bandwidth. Unfortunately, this cannot be quantitatively specified because we were unable to test disparities beyond $\pm 30$ arcmin. Because of the clearly asymmetric nature of these tuning functions, their bandwidths cannot be fairly characterized by expressing half-width at half-height as the halfwidth could only be measured in the direction of the steeper falloff. It is obvious, however, that these functions are asymmetric. It is also evident that the crossover from excitation to inhibition for these tuning functions occurs on the same side of the horopter as the peak of the tuning function. In other words, these tuning functions demonstrate that a stimulus with a crossed disparity can inhibit other stimuli with crossed disparities.

**Fig. 7.** Same as Fig. 6, but for subject LKC. The zero-disparity tuning function is even symmetric and has a narrower bandwidth than the other tuning functions. The off-horopter tuning functions tend to be odd symmetric and, as with the other subject, the negative-going (inhibitory) lobe positioned toward the horopter relative to the positive-going lobe.

**Fig. 8.** The family of tuning functions for subject LKC. This plot indicates the trend toward increasing bandwidth as a function of peak disparity. It also illustrates that the tuning functions do not necessarily cross from excitatory to inhibitory at the horopter. Rather, two surfaces can both be on the same side of the horopter and yet be mutually inhibitory.

**Discussion**

*Comparison with previous work from our laboratory*

Prior to this work, we conducted a study to investigate disparity-tuning functions by using a method of adaptation (Stevenson et al., 1992). Fig. 9 shows comparisons for the zero-disparity tuning functions and two of 10-arcmin tuning functions.

In this figure, the vertical position of the tuning functions is not a free parameter. A "log adaptation effect" of zero (indicating that the stimuli were processed by independent channels) must be equal to a summation index of zero (again, indicating that the stimuli were processed by independent channels). The relative vertical scale, however, was a free parameter and was chosen to yield the best visual match between the two tuning functions. As can be seen from the comparisons, the tuning functions derived from these different psychophysical techniques agree quite well in terms of peak location, bandwidth, crossover location, and symmetry.

A more extensive comparison of the tuning functions is shown in Fig. 10. In this figure, the heights of the tuning functions from the adaptation study were plotted as a function of the heights of the comparable tuning functions from the present study for all disparities at which the two tuning functions both had a data point. For example, from the lower panel of Fig. 9, we simply plot the height of the +10-arcmin data point from the adaptation tuning function against the height of the +10-arcmin data point from the present study. This yields a point at roughly (0.11, 0.29) on the graph in Fig. 10. In a sense, the scatter plot of Fig. 10 captures the coincidence of all of the features of the tuning functions, in that the correlation of the scatter plot would be disrupted by a dissimilarity in any of the features of the tuning functions (e.g. the bandwidth).

Thus, it seems that methods of both subthreshold summation and adaptation yield common tuning functions in the disparity
Fig. 9. Representative comparisons between the tuning functions obtained by subthreshold summation and those obtained by the adaptation paradigm of Stevenson et al. (1992). 10-arcmin crossed disparity, zero disparity, and 10-arcmin uncrossed disparity tuning functions are shown. The vertical scale (but not the vertical offset) was a free parameter, and was chosen to yield the best visual match.

Fig. 10. Scatter plot comparing the tuning functions derived with the subthreshold summation paradigm with those derived with the adaptation paradigm. The height of the adaptation tuning function (given by the ordinates) was plotted against the height of the comparable subthreshold summation tuning function. The data are collapsed across the two subjects and across the peak disparity of the tuning functions. The equation of the best-fit regression line is \( y = 1.2x + 0.07 \) with \( R^2 = 0.601 \).

domain, and are analogous to the situation that exists in color vision (i.e. the wavelength domain). This is more advantageous than the situation that exists in spatial vision, where the subthreshold summation paradigm yields much narrower tuning functions than those derived using other methodology, unless it is assumed that a type of "within mechanism probability summation" occurs, and the data are corrected for the effects thereof (Wilson & Bergen, 1979). Much could be hypothesized concerning the difference between the manner in which the spatial-frequency domain and the domains of disparity and wavelength are handled by the visual system. Since we do find equivalent tuning functions in the disparity domain, an extended discussion of the further assumptions required in the spatial-frequency domains will not be attempted here.

**General discussion**

Since the ground-breaking papers by Whitman Richards (1970, 1971, 1972) were published, very little attention has focused on the psychophysics of disparity channeling. While "motion-in-depth" channels, which involve changes in disparity over time, were revealed by Regan and Beverley (1978), the attention of most psychophysicists interested in stereopsis has been focused elsewhere. Primarily, psychophysical investigators have attempted to elucidate the spatial channeling characteristics of stereopsis (e.g. Yang & Blake, 1991; Heckman & Schor, 1989; Mayhew & Frisby, 1978; Julesz & Miller, 1975; Tyler, 1974).

Richards (1970, 1971, 1972) suggested that retinal disparity is encoded by the activity of pools of disparity detectors. Richards (1971) found that there existed classes of "stere anomalouss" observers, two of whose members were selectively insensitive to either crossed or uncrossed disparities, and a third whose members correctly perceived the magnitude of the disparities, but were often confused about its sign. Based on these results, he proposed that there "may be as few as three basic pools" of binocular activity. This explanation was very attractive due to its parsimony, coming as it did at a time when the vision community was beginning to realize that an elegant model based on a small number of spatial-frequency channels would probably not explain most psychophysical results in spatial vision.

Yet the three-pool hypothesis is problematic. A small number of channels along a dimension yield characteristically "bumpy" functions of both sensitivity and increment sensitivity along that dimension. For example, because spectral information is processed by three cone types, both the wavelength increment sensitivity (Wright & Pitt, 1934) and wavelength-specific luminance increment-sensitivity functions, when performed on a
suitable background (Stiles, 1949; Kranda & King-Smith, 1979; Thorton & Pugh, 1983), are composed of bumps and dips which are the signature of the underlying cone spectral sensitivities. However, no such signature is present across subjects in the disparity-increment threshold data of Badcock and Schor (1985), the correlation detection data of Stevenson et al. (1992), or the data of the present study.

Lehky and Sejnowski (1990) realized that the disparity-increment threshold data of Badcock and Schor (1985) were inconsistent with a system of coarse disparity pooling. This led them to develop a model of stereocuity based on a “distributed representation” of disparity. It should be noted that they used difference-of-Gaussian profiles for the tuning profiles of their disparity-tuned units, which is equivalent to incorporating disparity domain inhibition into the model. Briefly, they found that they were unable to model the data of Badcock and Schor (1985) by adhering to the near-far-tuned trichotomy of disparity processing; the resulting increment-sensitivity data displayed the frequent bumps and dips characteristic of systems comprising a small number of channels. The smallest number of units which Lehky and Sejnowski felt gave an adequate representation of the Badcock and Schor data was 17. These units were distributed across disparity with the peak sensitivity decreasing, and the bandwidths increasing, as a function of the peak disparity of the unit.

In modeling our disparity adaptation data (Stevenson et al., in press), we were forced to adopt a similar conclusion. We were unsuccessful in simulating tuning functions resembling those which we derived psychophysically with only three disparity-tuned channels. We found it possible to obtain an adequate fit to our data by using 21 units which, like those of Lehky and Sejnowski (1990), had a difference-of-Gaussian profile.

The models of Lehky and Sejnowski (1990) and Stevenson et al. (in press) is not at odds with other models of “stereopsis.” This is because, in general, modeling efforts have been focused on solving the correspondence problem and, in so doing, generating a disparity map of the stimulus.‡ Such a disparity map would serve as the input to a model such as that of Lehky and Sejnowski (1990).

There is another hallmark of systems comprising a small number of channels. Tuning functions derived for such a system will generally peak at a small number of locations, corresponding to the peaks of the underlying channels. Classic examples of this behavior are the W mechanisms of Stiles (1949). In a system comprising multiple overlapping channels, however, tuning functions will generally peak along a continuum of loci, as is observed in the spatial-frequency domain. This is also the behavior observed for tuning functions in the disparity domain. In addition to the data of the present study and the data of Stevenson et al. (in press) discussed above, the work of Felton et al. (1972) supports this notion. These authors measured contrast-detection thresholds as a function of disparity following adaptation to a high-contrast grating with some retinal disparity. Invariably, the biggest decrement in contrast sensitivity was found at the disparity of adaptation. In other words, the peaks of the tuning functions did not fall in a few discrete locations, but rather occurred wherever the initial adapting stimulus was positioned.

Recent physiological evidence also indicates to us that disparity is processed not by a three discrete disparity-tuned mechanisms, but rather by a larger, more distributed set. Poggio et al. (1988) tested the disparity sensitivity of cells in the visual cortex of the awake, behaving macaque. Previous physiology had grouped neurons into three categories, “tuned,” “near,” and “far,” corresponding to the original Richards trichotomy (Poggio & Fischer, 1977). Poggio et al. (1988), however, found it necessary to incorporate two new categories into their classification scheme, the “tuned near” and “tuned far” types, which represented neurons whose tuning properties were midway between those of the “tuned” type and those of the “near” and “far” types, respectively. Poggio et al. (1988) themselves hypothesize that “it may well be that there exists a continuum of disparity selectivity profiles from the extended near/far profiles asymmetric about zero disparity, to those with tuning peaks outside Panum’s region” (p. 4541). If Panum’s area is taken to be on the order of 6 arcmin wide for bar stimuli, then it would seem that all of the cells fell along a continuum of tuning peak location, as a cell was considered “tuned zero” if its tuning peak fell within ±3 arcmin of zero disparity. That is, it seems unlikely that much utility is to be gained in drawing a strict distinction between a cell which is tuned to 3 arcmin of crossed disparity and one which is tuned to 6 arcmin of crossed disparity (and which has a similar bandwidth).

It is perhaps worth mentioning that the tuning profiles which we derived psychophysically bore a strong resemblance to the “tuned zero,” “tuned near,” and “tuned far” tuning profiles published by Poggio et al. (1988). The bandwidths were quite similar, particularly between our zero disparity-tuning functions and the “tuned zero” tuning profiles of the single cells. In addition, both sets of data showed similar inhibitory flanks.

LeVay and Voight (1989) conducted an extensive study of the disparity tuning of cells in the visual cortex of the cat. The disparity-tuning profiles which they found for individual cells were quite similar in shape to those found in the macaque by Poggio’s group. LeVay and Voight (1989), however, specifically addressed the question of whether or not disparity-tuned cells in cortex fell into discrete categories or “types.” They conclude that “cortical cells display a broad continuum of disparity-tuning properties, both in terms of their sensitivity to disparity and in the shapes of their tuning curves.” Thus, it appears that the physiological evidence fails to support the hypothesis that disparity processing is subserved by a trichotomy of disparity-tuned units or channels.

Conclusions
Based on the above evidence, we propose that disparity processing is accomplished via a continuum of disparity-tuned chan-
nels. Rather than falling into discrete categories, we believe that these channels are distributed in terms of their tuning peaks and bandwidths. As we find strong evidence for disparity domain inhibition in our data, these channels must either be mutually inhibitory, or have a center-surround (e.g. difference-of-Gaussian) type profile.

References


Appendix

Given that the size of the effects we report in this manuscript are so large, the conclusion that disparity domain inhibition exists does not depend on the model we use to generate the null hypothesis (that no inhibition exists). However, as the concept of probability summation is central to our data analysis, it is discussed more fully below.

The probability of detection as given by the probability summation model for combining observations by independent channel is

\[ p(d) = 1 - \prod_{n=1}^{N} (1 - p_n(d)) \]

(1)
where $p_n$ is the probability of the $n$th channel detecting the stimulus. For two-alternative forced-choice, this can be readily converted to percent correct:

$$p_n(C) = 1 - \frac{1}{2} \prod_{n=1}^{N} (1 - p(d'))$$.

(2)

There are at least two other widely used models for combining information from two independent channels or observations. Given the size of the effects that we report, and the fact that we deal with only a two-channel analysis, it is not critical which model is adopted. For any number of channels below 10, the predictions of a vector-magnitude model and the probability summation model are trivial (Quick, 1974).

The other main alternative model is the $d'$ integration model (Green & Swets, 1966) in which

$$d' = \sqrt{\sum_{n=1}^{N} (d'_n)^2}$$.

(3)

To ensure that any differences between the predictions of the $d'$ integration model and the probability summation model would not affect our arguments, the following calculations were done.

We modeled the response functions in each of two hypothetical channels as Weibull functions:

$$p_n(d) = 1 - (1 - \gamma)e^{-[c_n/\alpha_n]^\beta}$$.

(4)

where $p_n(d)$ is the probability of detection by channel $n$, $c_n$ is the input signal strength, $\gamma$ is the lower asymptote, and $\alpha$ and $\beta$ are parameters determining the position and the slope of the response function, respectively.

Using 11 different signal strengths in each channel (i.e. 11² combinations), we calculated the probability of detection based on both the probability summation model and the $d'$ integration model.

For the probability summation model, the output of eqn. (4) to two input signals, $c_1$ and $c_2$, were used in eqn. (2). Similarly, percent correct values from eqn. (4) can be converted to $d'$ by using the inverse cumulative normal (cf. Elliott, 1964). These values were then used in eqn. (3), and the resulting values transformed back to percent correct to yield predictions of the $d'$ integration model which were directly comparable to the predictions of the probability summation model.

The predictions of the two models turn out to be virtually indistinguishable when the value of $\beta$ is near two. The resulting data can be plotted in a coordinate system in which the values of input signal strength are given by the $X$ and $Z$ coordinates and the proportion of correct responses is given by the height, $Y$, above the $X$-$Z$ plane. Iso-percent correct contours on the two surfaces (probability summation and $d'$ integration) are almost identical in form when $\beta = 2$, with the $d'$ integration-predictions deviating slightly from circular form.

In this paper, we used the unit circle as the prediction for probability summation. This is equivalent to using a vector sum of the individual (single observation) thresholds as the prediction for the combined (two observation) threshold. This can be written as

$$K_{AB} = c_A^2 + c_B^2$$.

where $K_{AB}$ is a constant and $c_A$ and $c_B$ are the single observation thresholds, respectively.

Finding a threshold for various combinations of $c_A$ and $c_B$ is equivalent to finding values of $c_A$ and $c_B$ for which probability of detection is constant. Thus, setting $p(d')$ equal to a Weibull function to represent the probability for single-channel detection, we have

$$k = 1 - (1 - W(c_A))(1 - W(c_B))$$.

(6)

where $W(c)$ is shorthand for eqn. (4). Substituting eqn. (4) with $\gamma = 0$ (the probability for single-channel detection) into eqn. (6) and rearranging, we have

$$e^{[c_A/\alpha_A]^{\beta_A} + [c_B/\alpha_B]^{\beta_B}} = \frac{1}{1 - k}$$.

(7)

since we are normalizing the data so that single observation threshold $= 1.0$ and assuming $\beta_A = \beta_B, \alpha_A = \alpha_B$ also. By taking the natural logarithm of both sides of eqn. (7) and isolating all of the constant terms on the right and representing them by a single constant, we have

$$c_\beta^A + c_\beta^B = K$$.

(8)

Obviously for $\beta = 2$, eqn. (8) is equivalent to eqn. (5) and is the equation of a circle. Letting $\beta = 2$, two special cases are worth comparing. When the signal strength is equal for the two observations, $c_A = c_B = c$, and

$$c = \frac{\sqrt{K}}{\sqrt{2}}$$.

(9)

When there is only one observation with signal strength $c_x$

$$c_x = \sqrt{K}$$.

(10)

combining eqns. (9) and (10) gives

$$c_x = c\sqrt{2}$$.

(11)

which is the often quoted "root two improvement" expected from probability summation.