

# Spatial and temporal coherence in perceptual binding

RANDOLPH BLAKE\* AND YUEDE YANG

Department of Psychology and Vanderbilt Vision Research Center, Vanderbilt University, Nashville TN, 37240

Communicated by George Sperling, University of California, Irvine, CA, April 23, 1997 (received for review July 1, 1996)

**ABSTRACT** Component visual features of objects are registered by distributed patterns of activity among neurons comprising multiple pathways and visual areas. How these distributed patterns of activity give rise to unified representations of objects remains unresolved, although one recent, controversial view posits temporal coherence of neural activity as a binding agent. Motivated by the possible role of temporal coherence in feature binding, we devised a novel psychophysical task that requires the detection of temporal coherence among features comprising complex visual images. Results show that human observers can more easily detect synchronized patterns of temporal contrast modulation within hybrid visual images composed of two components when those components are drawn from the same original picture. Evidently, time-varying changes within spatially coherent features produce more salient neural signals.

Early vision entails local feature analyses of the retinal image carried out in parallel over the entire visual field. By virtue of the receptive-field properties of the neurons performing this analysis, visual information is registered at multiple spatial scales, ranging from coarse to fine, for different contour orientations (1–3). Moreover, different qualitative aspects of the visual scene engage populations of neurons distributed among numerous, distinct visual areas (4, 5). Yet we perceive objects whose constituent features are, at least metaphorically speaking, bound together coherently. One popular but controversial hypothesis posits as a binding agent temporal synchronization of neural activity among cortical cells registering object features (6–11). In experiments reported here, we have discovered that synchronized modulations over time in the contrast of separate components of complex images are easier to detect when those components form a meaningful object. These findings are compatible with the notion that temporal and spatial coherence are involved in the promotion of perceptual binding.

Several recent experiments have tried to assess whether patterns of neural activity coincident in time promote perceptual grouping, by determining whether visual features flickering in temporal synchrony more readily promote figure/ground segregation. Results from those experiments, however, have led to contradictory conclusions (12–14). To pursue this question of figural binding from a complementary perspective, we tested for enhanced detectability of temporal synchrony among spatial features that define a visual object. To understand the rationale for our study, imagine a picture composed of two components each selected to activate separate populations of visual neurons. Suppose further that the contrast of each component can be independently varied over time (Fig. 1), with the temporal pattern of contrast modulations of the components being either identical (i.e., synchronized) or uncorrelated (i.e., unsynchronized). Physiological work (16–18) shows that fluctuations in contrast amplitude over time will

evoke corresponding temporal modulations in neural activity. Therefore, discrimination of synchronized from unsynchronized contrast modulations of the two components would depend on information contained in the temporal patterning of activity within the separate populations of neurons responsive to those components. Is the ability to detect synchronized patterns of temporal modulation easier when the two components form a single object? If spatially coherent features more readily generate temporally correlated neural activity (15), the answer should be “yes.”

## Temporal Modulation at Multiple Spatial Scales

Experiment 1 examined image components represented at different spatial scales (Fig. 2A). Gray scale images were spatial-frequency filtered (19, 20) into “low-pass” (LP) and “high-pass” (HP) components designed to promote activation of separate neural mechanisms (1–3). Hybrid images were then created by combining a LP image and a HP image, with members of each pair drawn either from the same original image or from different originals. The contrast of each component of a pair was modulated in small, random steps over time, with the pattern of random contrast steps independently specified for each component. Observers viewed two successive presentations of these hybrid images (Fig. 1). During one presentation contrast modulations were identical for the two components (synchronized), and during the other presentation contrast modulations were uncorrelated (unsynchronized). Over different blocks of trials, the rate of contrast modulation was varied, holding exposure duration constant. Observers indicated in which interval the contrast modulations were synchronized, without regard for the identity of the components. To reiterate, the information necessary for performing this task is contained in the temporal patterns of neural activity within the two populations of neurons activated by the HP and the LP components.

Synchronized contrast modulations were detectable within smaller contrast ranges when the two components were drawn from the same original, compared with conditions where the two were drawn from different originals (Fig. 2B), particularly at higher rates of temporal modulation. Moreover, superior detection performance was most pronounced when the LP and HP components were drawn from an original picture of a human face. Performance was poorest when both filtered components were drawn from a random dot pattern, with performance measured with images drawn from natural scenes and nonhuman objects being intermediate. An ancillary experiment established that the advantage of face images was obtained only when both HP and LP components were drawn from the same face and when both were presented in the same orientation. Evidently the equivalence in global structure (technically speaking, the equivalent phase spectra) between LP and HP images drawn from the same original more readily supports the detection of spatio-temporal coherence of those images.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1997 by The National Academy of Sciences 0027-8424/97/947115-5\$2.00/0

Abbreviations: LP, low pass; HP, high pass.

\*To whom reprint requests should be addressed. e-mail: blaker@ctrvax.vanderbilt.edu.

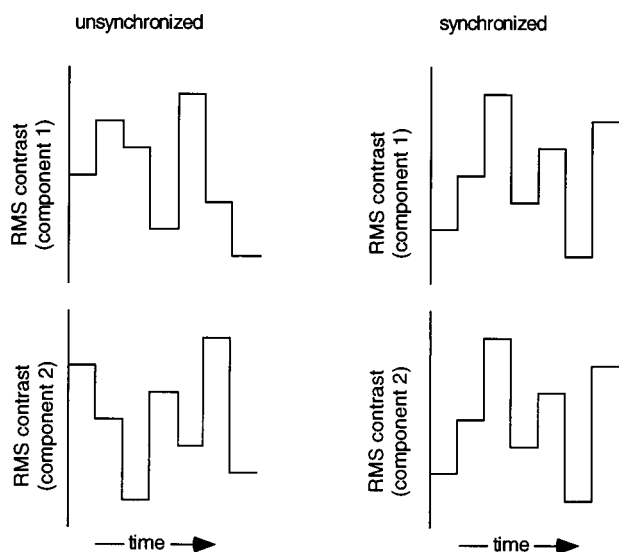


FIG. 1. Hybrid images composed of two components were produced using techniques summarized in Figs. 2A and 3A. Regardless of hybrid type, the contrast of each of the two components could be varied over time, with the temporal pattern of contrast modulations being either identical for the two components (synchronized) or uncorrelated for the two (unsynchronized). For each stimulus presentation, seven different contrast values were presented in immediate succession (no blank interval), with each successive contrast value selected at random without replacement. These contrast modulations were always centered about a value of 0.3 rms, the step-size between contrast values was constant in log units and the contrast range (maximum–minimum contrast values in the temporal sequence) was varied to manipulate the discriminability of synchronized from unsynchronized presentations. Discrimination was progressively more difficult at smaller contrast ranges. Different rates of contrast modulation could be achieved by manipulating the frame duration (where frame refers to presentation of a given contrast level).

Are observers simply picking the display interval in which the HP and LP components appeared different in contrast, without regard for their temporal pattern of contrast modulations? This possibility is ruled out by results from a control condition in which observers viewed hybrids composed of HP and LP components whose contrast values remained unchanged during the 580-msec presentation. All four experienced observers performed at chance levels when required to discriminate hybrids in which the two superimposed, static components differed in contrast by as much as 0.27 log units (a value equivalent to the largest contrast difference between components associated with the threshold values shown in Fig. 2B) from ones in which the static components were equal in contrast. The inability to perform this control task underscores that observers indeed relied on temporal modulations in contrast to perform the original task.

In another control condition, we determined whether a one-frame temporal phase shift of one component's contrast modulation would disrupt the detection of spatio-temporal coherence. On each trial of this experiment, observers viewed two successive intervals: one in which the contrast modulations of the HP and LP components were unsynchronized (i.e., random with respect to one another) and the other in which both components followed the same pattern of modulations except that the modulation steps for one component were offset in time by one frame relative to the modulation steps in the other component (with the last frame of this delayed sequence wrapped around to the first position in the sequence). Thus in these presentation intervals, the pattern of contrast modulations was identical for the two components, but the two sequences were offset in time

and, hence, asynchronous. Observers were instructed to pick the interval in which the pattern of modulations was identical but phase shifted. Two practiced observers were tested at 24- and 36-Hz modulation rates, values at which performance for the "same" hybrids was significantly better than for the "different" hybrids in our main experiment. Performance on this phase shift condition never exceeded chance, even for the very largest contrast range where performance was essentially perfect in the main experiment. So regardless whether the hybrids were composed of same or different components, the phase lag destroyed temporal coherence thus rendering the task impossible, as predicted if performance is mediated by detection of synchronous modulation of neural activity.

Can synchronized contrast modulation of LP and HP components drawn from different originals promote "false" binding of those components? Our task does not address this question, but observers offered revealing comments. Hybrids composed of synchronized but dissimilar components created an impression of one component transparently in front of the other. Unsynchronized components, rather than creating transparency, seemed to compete for attention.

### Temporal Modulation of Spatially Distinct Features

Experiment 2 measured the detectability of temporal coherence between images whose components were separated in space, not in spatial frequency (Fig. 3A). Unfiltered pictures were cut in half, yielding upper and lower components. Hybrids were then generated using these components, with half of the hybrids consisting of components drawn from the same original and half consisting of components drawn from different originals. The upper and lower portions of the hybrids were separated by a blurred, horizontal strip whose width was 25, 45, or 65 min arc and whose uniform luminance was equivalent to the space-average luminance of the upper and lower halves of the image. The contrast of the upper and lower components was modulated synchronously during one test interval and asynchronously during the other interval, with the observer's task being to indicate in which interval contrast modulations were synchronous. Each of the seven consecutive frames of a given presentation was 83.3 msec in duration, corresponding to a modulation rate of 12 Hz. The contrast range, which was always centered about an rms value of 0.30, was varied randomly over trials to find the range associated with 75% correct performance.

Observers were more accurate at detecting synchronized contrast modulations when the two halves belonged to the same original (Fig. 3B), but only when those two components were presented in exact synchrony—introduction of a temporal phase lag between otherwise identical patterns of contrast modulation transformed an easy task into an impossible one. Enhanced detectability of synchronization of same components was measurable over spatial extents in excess of 1° for these centrally fixated images. This finding, too, makes sense in terms of a temporal binding mechanism that operates to collate features of a given object over space, including viewing conditions where central portions of the object are occluded (21). Results from experiment 2 also complement a recent study demonstrating that texture features defining a figure are more easily segregated from their background when those figural features appear and disappear in synchrony (14). Evidently, temporal synchrony is not absolutely necessary for figural binding, however, because disruptions in the temporal synchrony of figural elements does not inevitably impair perceptual grouping (12, 13).

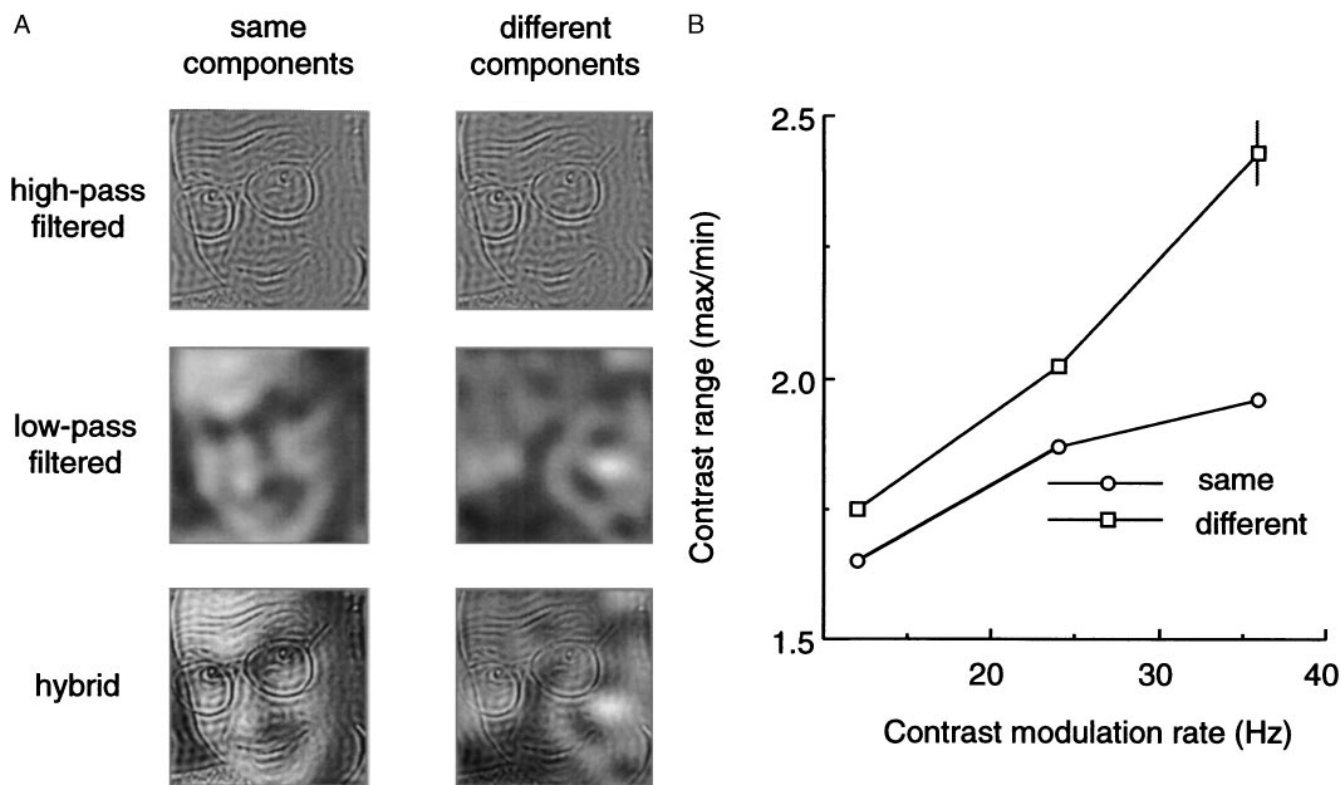


FIG. 2. (A) Starting with 8-bit gray scale images of natural scenes, faces, and dot patterns, hybrid pictures were created by spatial frequency filtering. Original pictures were filtered to produce HP images (all low spatial frequencies removed) and LP images (all high spatial frequencies removed); LP and HP filter cutoff values (3 dB) were separated by 1.5 octaves and, at the 1.07 m viewing distance, corresponded to 2.35 c/deg and 6.57 c/deg; these LP and HP values were selected to promote activation of separate populations of spatially frequency-tuned neurons. Various HP and LP components were then combined (i.e., gray scale values added on a pixel-by-pixel basis) to produce hybrid pictures in which the two components were drawn from the "same" original (e.g., LP and HP components from a given face) or were drawn from "different" originals (e.g., HP from a face and LP from random dots). A given hybrid image was presented during both intervals of a two-interval, forced-choice trial; during one, randomly selected interval the pattern of temporal contrast modulations of the two hybrid components was identical (i.e., synchronized) and during the other interval the modulations were uncorrelated (i.e., unsynchronized). On half of the trials the LP component and the HP component of the hybrid were drawn from the same original (i.e., same condition), and on the remaining half of the trials the LP and HP components were drawn from different originals (i.e., different condition); "same" and "different" trials were randomly intermixed within a block of 60 trials. On each trial, observers simply indicated in which interval the contrast modulations were synchronized, guessing if necessary without feedback. On any given trial, contrast modulations in the two successive presentations were restricted to a range of contrast values centered around an rms value of 0.30; the range was identical on both intervals of any trial, but the range varied randomly from trial to trial. Trials were presented as a method of constant stimuli, to determine the contrast range where observers were able to discriminate synchronized from unsynchronized modulations 75% of the time based on probit analysis. Over different blocks of trials, the rate of contrast modulation was either 12, 24, or 36 Hz (corresponding to frame durations of 83.3, 41.6, or 27.7 msec). In one condition, exposure duration remained constant at 583 msec (such that the total number of frames presented varied directly with modulation frequency), and in another condition the number of frames presented was always seven (such that exposure duration varied inversely with modulation frequency); both of these conditions yielded the same pattern of results for all four observers. The gray scale pictures (2.1 deg<sup>2</sup>; 24.5 cd m<sup>-2</sup> average luminance) were presented on a Radius video monitor (1152H × 882V pixel resolution; P104 phosphor; 72 Hz refresh rate) under control of an accelerated Macintosh IIX computer. Calibrated look-up tables corrected luminance nonlinearities. Observers initiated trials and indicated responses using keys on the computer keyboard; error feedback was not given except during a block of 10 practice trials preceding each test session. Each of four observers completed 1,200 trials for each type of stimulus. (B) Threshold contrast range (expressed as the ratio of the maximum and minimum contrast values) for discriminating synchronized from unsynchronized contrast modulations when HP and LP components were drawn from the same original (Same) and when those components were drawn from different originals (Different); average results for the four observers are shown for each of the three modulation rates tested under the condition where trials always involved seven-frame presentations (i.e., exposure duration varied inversely with modulation rate). The pattern of results was identical for all four observers, and the differences between same and different conditions are statistically significant (z-score two-tailed test,  $P < 0.001$ ); the vertical bar denotes the average standard error. The observer's task had nothing to do with identifying the components of the hybrids, only whether or not the temporal pattern of modulations in contrast of the two components was identical. Threshold values were impossible to estimate under a condition where the two components of the synchronized condition were offset by one frame, because these phase lag presentations were always indiscriminable from the unsynchronized presentations regardless of contrast range.

## Conclusions

Our results demonstrate that it is easier to detect coincident changes in component features over time if those features together constitute a meaningful object. From this we conclude that time-varying changes within spatially coherent features produce more salient neural signals (15), although it remains to be learned just what constitutes "spatial coherence" for purposes of this task. Our conclusion is

consistent with models of binding in which extrinsically activated neural mechanisms resonate to spatio-temporal coherence among local features comprising a global object or event (6, 10). It has also been proposed that visual features lacking externally imposed temporal structure can be grouped by virtue of internal generation of synchronized activity (22, 23), although that idea is controversial (24, 25). We stress that our data have no direct bearing on the efficacy of internally generated neural oscillations, although it is

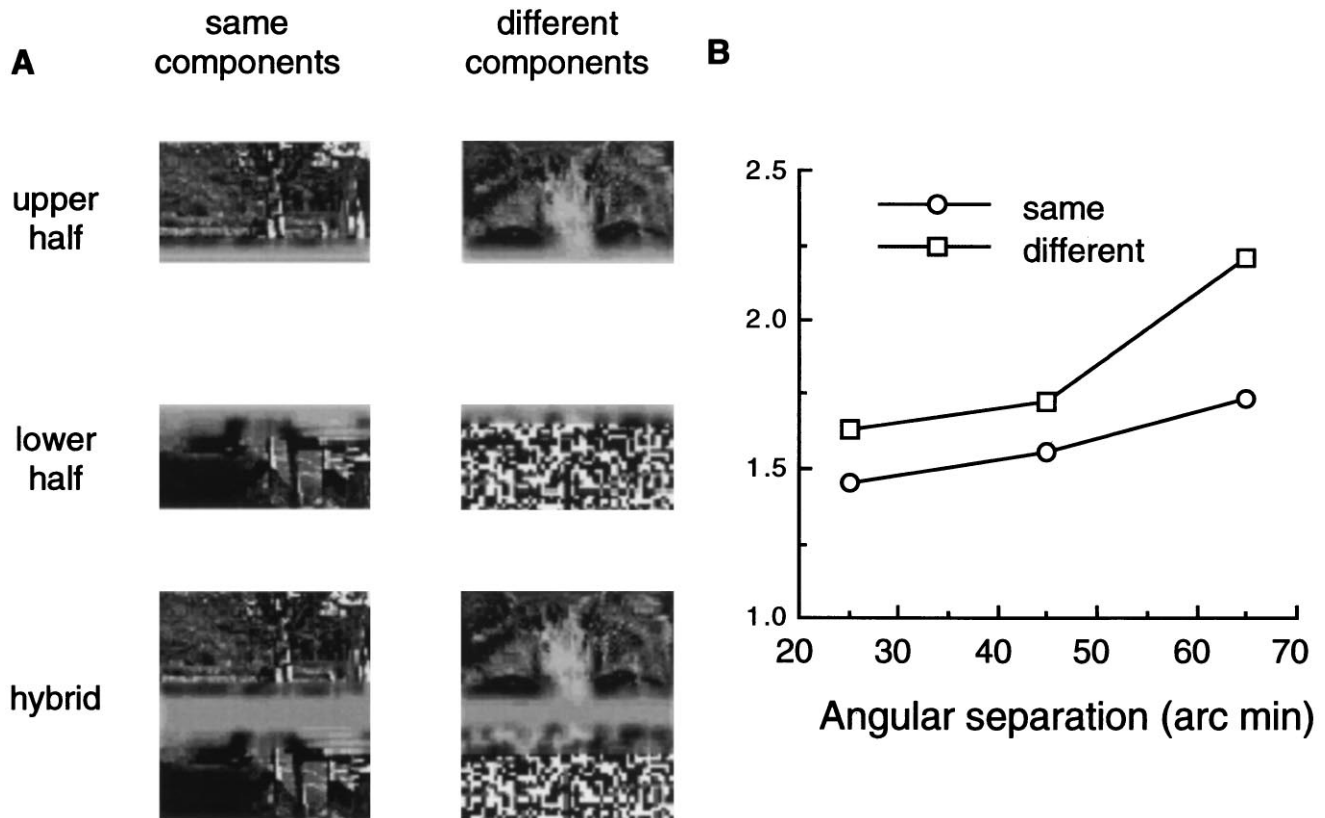


FIG. 3. (A) Examples of unfiltered pictures that were cut in half, with the upper and lower components then recombined to produce hybrids consisting of components drawn from the same original and hybrids consisting of components drawn from different originals. (B) Threshold contrast range for discriminating synchronized from unsynchronized contrast modulations for unfiltered pictures in which the upper and lower portions were drawn either from the same original (Same) or different originals (Different). The abscissa plots the size of the blank gap separating upper and lower portions of the hybrid picture; data are averaged over four observers and the difference between same and different conditions is statistically significant (z-score two-tailed test,  $P < 0.001$ ).

certainly possible that such oscillations provide the carrier signal for extrinsically triggered, synchronized activity.

While consonant with the notion of temporal coherence as a neural binding agent, our findings with HP and LP components pose a conundrum concerning the arrival of neural information at the cortical level. Physiological recordings indicate that afferent signals carried by pathways maximally responsive to low spatial frequencies activate cortical neurons about 20 msec sooner than do signals from pathways responsive to high spatial frequencies (26, 27). Yet our data reveal no hint of a temporal phase lag of the HP component relative to the LP component. How the brain reconciles these latency differences arising in the retina remains an intriguing question, although it may have something to do with the very rapid, dynamic organization of neural activity arising among spatially neighboring neurons (28).

Our technique and findings set the stage for further investigations concerning visual binding. For example, what constitutes spatial structure from the standpoint of detection of temporal coherence? Does temporal synchrony make it easier to judge whether components come from the same vs. different originals? Do these robust effects generalize to other, more ecologically valid forms of temporal modulation such as object motion? Finally, by demonstrating that the visual system is especially sensitive to spatio-temporal coherence, our novel procedure offers a promising avenue for studying the binding problem neurophysiologically.

We thank George Sperling, Jeffrey Schall, David Gilden, Ute Leonards, Wolf Singer, and Vicki Ahlstrom for helpful discussion. This work was supported by National Institutes of Health Grants EY07760 and EY01826.

- Marr, D. (1982) *Vision* (Freeman, San Francisco).
- Wandell, B. (1995) *Foundations of Vision* (Sinauer, Sunderland, MA).
- DeValois, R. & DeValois, K. (1988) *Spatial Vision* (Oxford, New York).
- Van Essen, D. C., Anderson, C. H. & Felleman, D. J. (1992) *Science* **255**, 419–423.
- Zeki, S. (1993) *A Vision of the Brain* (Blackwell, Cambridge, MA).
- Damasio, A. R. (1989) *Neural Comput.* **1**, 123–132.
- Crick, F. & Koch, C. (1990) *Semin. Neurosci.* **2**, 263–275.
- Hardcastle, V. (1994) *J. Consciousn. Studies* **1**, 66–90.
- Singer, W. (1990) *Concepts Neurosci.* **1**, 1–26.
- Milner, P. M. (1974) *Psychol. Rev.* **81**, 521–535.
- von der Malsburg, C. (1986) in *Brain Theory*, eds. Palm, G. & Aertsen, A. (Springer, Berlin).
- Fahle, M. & Koch, C. (1995) *Vision Res.* **35**, 491–494.
- Kiper, D. C., Gegenfurtner, K. R. & Movshon, J. A. (1996) *Vision Res.* **36**, 539–544.
- Leonards, U., Singer, W. & Fahle, M. (1997) *Vision Res.*, in press.
- Kreiter, A. K. & Singer, W. (1996) *J. Neurosci.* **16**, 2381–2396.
- Bodis-Wollner, I., Hendley, C. D. & Kulikowski, J. J. (1972) *Perception* **1**, 341–349.
- Troy, J. B. & Enroth-Cugell, C. (1993) *Exp. Brain Res.* **93**, 383–390.
- Hamilton, D. B., Albrecht, D. G. & Geisler, W. S. (1989) *Vision Res.* **29**, 1285–1308.
- Stromeyer, C., III, & Julesz, B. (1972) *J. Opt. Soc. Am.* **62**, 1221–1232.
- Yang, Y. & Blake, R. (1991) *Vision Res.* **31**, 1177–1190.
- Gilbert, C. D. (1992) *Neuron* **9**, 1–13.
- Singer, W. (1995) *Annu. Rev. Neurosci.* **18**, 555–586.
- Engel, A. K., Konig, P., Kreiter, A. K. & Singer, W. (1991) *Science* **252**, 1177–1179.

24. Ghose, G. M. & Freeman, R. D. (1992) *J. Neurophysiol.* **68**, 1558–1574.
25. Shadlen, M. N. & Newsome, W. T. (1994) *Curr. Opin. Neurobiol.* **4**, 569–579.
26. Munk, M. H. J., Nowak, L. G., Girard, P., Choulamountri, N. & Bullier, J. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 988–992.
27. Maunsell, J. H. R. & Gibson, J. R. (1995) *J. Neurophysiol.* **68**, 1332–1344.
28. Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H. & Aertsen, A. (1995) *Nature (London)* **373**, 515–518.