17 Jones, J. and Malik, J. (1992) A computational framework for determining stereo correspondence from a set of linear spatial filters *Proc. Eur. Conference on Computer Vision*, Genova, Italy, pp. 395–410

18 Anderson, B.L. and Nakayama, K. (1994) Towards a general theory of stereopsis: binocular matching, occluding contours, and fusion *Psychol. Rev.* 101, 414–445

19 Anderson, B.L. (1994) The role of partial occlusion in stereopsis *Nature* 367, 365–368

20 Anderson, B.L. and Julesz, B. (1995) A theoretical analysis of illusory contour formation in stereopsis *Psychol. Rev.* 102, 705–743

21 Anderson, B.L. (1997) A theory of illusory lightness and transparency in monocular and binocular images: the role of contour junctions *Perception* 26, 419–453

22 von Szilzy, A. (1921) *Graefes Arch. Ophthalmol.* 105, 964–972

23 Nakayama, K. (1996) Binocular visual surface perception *Proc. Natl. Acad. Sci. U. S. A.* 93, 634–639

24 Julesz, B. (1964) Binocular depth perception without familiarity cues *Science* 145, 356–362

25 Julesz, B. and Chang, J. (1976) Interaction between pools of binocular disparity detectors tuned to different disparities *Biol. Cybern.* 22, 107–119

26 Ramachandran, V.S. and Cavanagh, P. (1985) Subjective contours capture stereopsis *Nature* 317, 527–530

27 Jordan, J.R., Geisler, W.S. and Bovik, A.C. (1990) Color as a source of information in the stereo correspondence problem *Vis. Res.* 30, 1955–1970

28 Smallman, H.S. and McKee, S.P. (1995) A contrast ratio constraint on stereo matching *Proc. R. Soc. London Ser. B* 260, 265–271

29 Rogers, B.J. and Bradshaw, M.F. (1993) Vertical disparities, differential perspective and binocular stereopsis *Nature* 361, 253–255

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

31 Gilchrist, A.L. (1994) Absolute versus relative theories of lightness perception, in *Lightness, Brightness and Transparency* (Gilchrist, A.L., ed.), pp. 1–34, Erlbaum

32 Nakayama, K., Shimojo, S. and Ramachandran, V.S. (1990) Transparency: relation to depth, subjective contours, luminance, and neon color spreading *Perception* 19, 497–513

33 Metelli, F. (1970) An algebraic development of the theory of perceptual transparency *Ergonomics* 13, 59–66

34 Metelli, F. (1974) Achromatic color conditions in the perception of transparency, in *Perception: Essays in Honor of J.J. Gibson* (MacLeod, R.B. and Pick, H.L., eds), pp. 96–116, Cornell University Press

35 Metelli, F. (1974) The perception of transparency *Sci. Am.* 230, 90–98

36 Metelli, F., da Pos, O. and Cavedon, A. (1985) Balanced and unbalanced, complete and partial transparency *Percept. Psychophys.* 38, 354–366

37 Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston

---

Perception of three-dimensional structure from motion

Richard A. Andersen and David C. Bradley

The ability to perceive the 3-D shape of objects solely from motion cues is referred to as structure-from-motion perception. Recent experiments indicate how this remarkable perceptual attribute is computed by the brains of primates. This computation proceeds in at least two stages, one in which motion measurements are made and another in which moving surfaces are reconstructed. The middle temporal area (MT) in the macaque monkey appears to play a pivotal role in the latter step and suggests a previously unappreciated function for this well-known cortical region, which had previously been thought to play a more rudimentary role in simply signaling the direction of motion of images.

R.A. Andersen and D.C. Bradley are at the California Institute of Technology, Division of Biology 216-76, Pasadena, CA 91125, USA.

tel: +1 626 395 8336
fax: +1 626 795 2397
e-mail: andersen@vis.caltech.edu, dbradley@vis.caltech.edu

Structure-from-motion (SFM) is the perception of depth induced by retinal motion. Although primates use other cues for depth perception, most notably stereopsis, the ability to perceive depth from motion is perhaps the earliest form of depth perception, and is believed to be used by lower animals such as insects and fish that lack stereopsis.

SFM, also known as kinetic depth, was studied psychophysically by Wallach and other psychologists in the early part of this century[1,2]. However, until recently, the neural mechanisms responsible for this amazing perceptual phenomenon have not been understood. We review here recent experiments that are identifying the anatomical pathways and neural algorithms responsible for three-dimensional (3-D) SFM perception.

Position versus motion cues

There are two general theories of SFM perception: those that use motion measurements[3–7] and those based on position measurements[8–12]. Perhaps the best known position-based model is Ullman's original incremental-rigidity algorithm[8]. This algorithm keeps track of the positions of specific points, across sample frames of the 2-D image, to reach a rigid, 3-D interpretation. Motion-based algorithms

PII: S1364-6613(98)01181-4

Box 1. Demonstrations of 3-D structure-from-motion

A dramatic demonstration of the SFM effect requires only an overhead projector and a paper clip. If one bends the paper clip into a complicated shape, and lays it on the surface of the overhead projector, the image projected is a two-dimensional (2-D) reflection of the twisted wire. This 2-D projection gives no indication of the three dimensional shape of the paper clip figure. However, if one rotates the paper clip under the light, the moving 2-D image is immediately perceived in 3-D.

A typical laboratory display used for SFM experiments is shown in Fig. A. Observers view the flattened image (2-D projection) of a revolving, random-dot cylinder. This display contains two sets of dots moving in opposite directions, and the speed of the dots varies sinusoidally with position within the stimulus, with the fastest speeds in the middle. Such a motion display is perceived as a 3-D rotating cylinder. However, this is also an ambiguous display because it does not specify which surface is in front and which is in back. For example, dots going right might initially appear to be in front, with the left-going dots in back, which translates to a cylinder that appears to be turning counterclockwise (from the top). But after a matter of seconds, the perceived surface order, and thus the perceived rotation, tends to flip spontaneously. This kind of SFM perception is said to be 'bistable', much like a Necker cube, whose perceived 3-D structure tends to invert spontaneously (see Fig. B). In the main text we review how the bistable property of the rotating cylinder display has been used to study the cortical areas and neural mechanisms responsible for SFM perception.

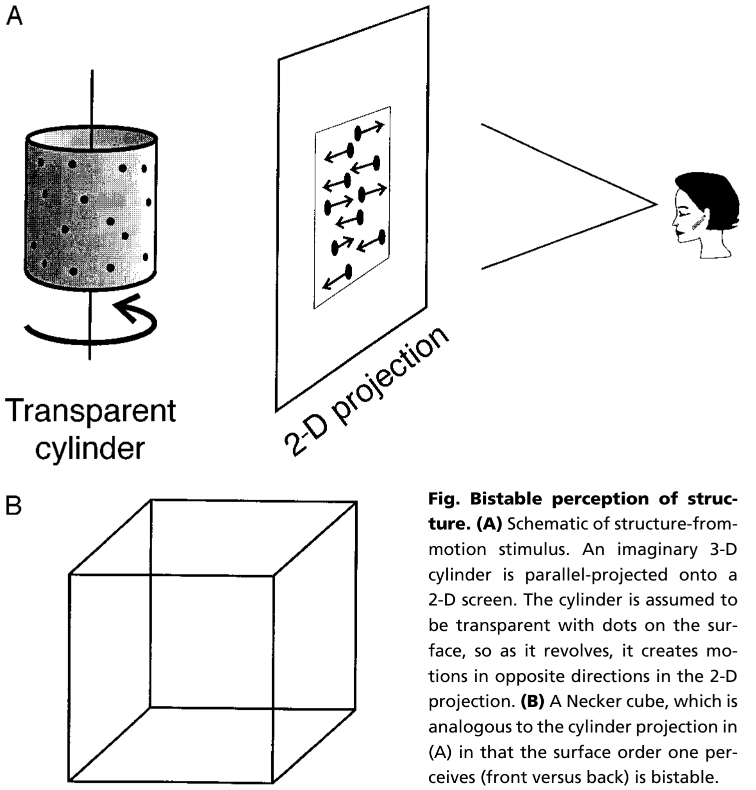

Fig. Bistable perception of structure. (A) Schematic of structure-from-motion stimulus. An imaginary 3-D cylinder is parallel-projected onto a 2-D screen. The cylinder is assumed to be transparent with dots on the surface, so as it revolves, it creates motions in opposite directions in the 2-D projection. **(B)** A Necker cube, which is analogous to the cylinder projection in (A) in that the surface order one perceives (front versus back) is bistable.

instead measure the local velocities of points on an image and use the global velocity field to compute 3-D SFM.

There are several lines of evidence suggesting that the brain uses motion information to compute SFM. Human psychophysical experiments demonstrate recovery of SFM in displays where the individual point lifetimes are only 80–100 ms (Refs 6,13–15). This minimum point lifetime is similar to the minimum lifetime required to make 2-D velocity judgments[16,17]. In position-based algorithms, performance is best for large displacements of the object between discrete image frames. Treue *et al.*[15] found that points must be visible for a minimum time, rather than change by a minimum displacement, for structure to be perceived. This result suggests that motion, and not position, is the critical measure. Also, lesions of MT produce deficits in monkeys' ability to perform SFM perceptual tasks[18]. Since MT is an area of the brain important for analyzing motion stimuli[19–21], this finding further supports the use of motion measurements in SFM perception. The motion measurements used for SFM perception may be encoded in the form of image velocity or, more simply, in the form of motion energy[13,14].

Surface interpolation

In some circumstances, such as when very dense dot displays are used, SFM can be perceived with the presentation of just two frames[22]. However, with a reasonable number of dots and limited point lifetimes (100–200 ms), vivid 3-D shapes like the cylinder shown in Box 1 can be perceived, but perception requires many frames over a long presentation time (600–1000 ms). This long processing time can be used to investigate how the 3-D percept is constructed by the brain. By using point lifetimes that are much shorter than the amount of time required to perceive structure, we have found that a surface interpolation mechanism is used to build up the 3-D percept[6,15]. In these displays, moving dots appear briefly and asynchronously at different locations, and this information is integrated over time to improve perception. Furthermore, if a small number of dots are presented over and over again in the same locations, subjects perceive little or no structure, whereas if the same number of points are replotted in new locations, the 3-D percept emerges. Thus, 3-D SFM perception requires moving points which are presented at a number of spatial locations, and this information can be gathered by using either a large number of dots, or a small number of dots that appear at different locations over time. These results provide strong support for the idea that the brain uses a surface interpolation mechanism, in which a mental representation of surfaces is constructed and updated over time.

Using the two ideas of velocity measurement and surface reconstruction, we were able to modify the original Ullman algorithm to account for the experimental observations mentioned above. In this new formulation[7], velocity measurements are made, and a 3-D surface is calculated from the available data. Then a new set of velocity measurements is compared with the existing surface, which is modified to take into account these new data. This process is repeated until an accurate percept is achieved.

This model, and particularly the surface component, explains a wide variety of SFM phenomena and illusions. It

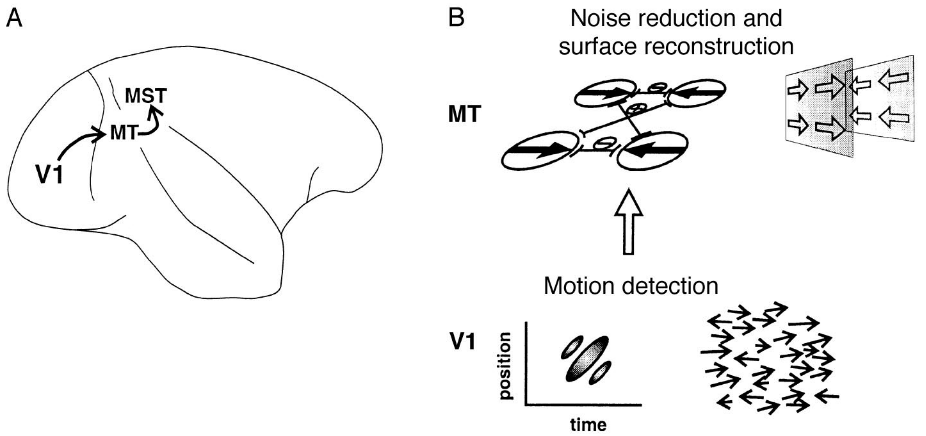

Fig. 1 Cortical motion-processing pathway in the primate. (A) Schematic diagram of the cortical motion pathway in the macaque brain. Many neurons in layer 4B of V1 (primary visual cortex) are direction-selective, and these project to the middle temporal area (MT) in the superior temporal sulcus, where nearly all neurons are direction-selective. MT, in turn, projects to the medial superior temporal area (MST), which contains direction-selective cells as well as cells (primarily in the dorsal portion, MSTd) tuned for complex motion patterns such as expansion and rotation. Consistent with this anatomical hierarchy, receptive-field diameters (typically <2 deg in V1, 5–20 deg in MT, and often >100 deg in MST) progressively increase, suggesting an integration and elaboration of information as one ascends the pathway. **(B)** Hypothetical roles of V1 and MT in motion processing. Local velocity (direction and speed) is initially measured in V1. Cell receptive fields are oriented in space and time (bottom left), acting as filters that selectively pass motion signals with a particular direction and speed. These local velocity estimates are combined in MT, where coherent motion signals are pooled to create a representation of moving surfaces, and opposing signals from a given depth (i.e. surface) are made to cancel, thus removing 'motion noise' due to flicker, etc. The output of MT neurons may in turn be combined in MST (not shown) to create receptive fields selective for optic flow patterns such as expansion and rotation.

also explains how the visual system solves specific and interesting problems. For instance, this model, like the visual system, is able to interpolate across areas within a surface where there are few features. Also, under transparent conditions or at motion borders, there are often motions in different directions which should not be averaged together, but rather assigned to separate surfaces; this is also accomplished by the model.

Anatomy of the motion pathway

In the hierarchy of primate visual processing, direction-selective cells first appear at the level of the primary visual cortex (V1) (Ref. 23; and see Fig. 1). These direction-selective cells are mainly confined to cortical layers 4B and 6 and project to the middle temporal area (MT) (Ref. 24). In MT, nearly all cells are direction selective[25], and they have much larger receptive fields than the V1 direction-selective neurons. Area MT projects robustly to a third area, the medial superior temporal area (MST). The dorsal aspect of this area (MSTd) has very large receptive fields, often covering much of the visual field, and the cells are often selective for particular patterns of motion such as expansion or rotation[26–30].

Our formal model, mentioned above, has two main stages which may map onto the primate cortical motion pathway (see Fig. 1). Of course, each stage of the model may be accomplished by more than a single cortical level, and we do not imply that there is a literal one-to-one mapping of the model onto cortical areas. However, we do wish to emphasize the hierarchical structure that exists both in the model and correspondingly in the motion pathway. The first stage is the measurement of motion signals, likely to occur in V1. The second stage, which probably includes MT, involves the reconstruction of surfaces. The large receptive fields in MT are conducive to integrating motion signals over space to construct surfaces. An additional aspect of this reconstruction stage assigns depth values across the surface through the use of motion gradient information. This 3-D surface computation may operate within MT, or may utilize the cells in MST which have been previously shown to be selective to various gradients and patterns of motion[26–30].

By using transparent motions, we have been able to establish that area MT is intimately involved in the surface-reconstruction process, and in particular in the removal of noise during this reconstruction process. Using bistable cylinders, we have also demonstrated that MT is involved in the depth ordering of surfaces. This segmentation by depth is important in surface reconstruction for determining which velocity measurements go with which surface, and is also essential for the overall 3-D percept itself. Discussions of these lines of research follow, each converging on the idea that area MT is crucial in 3-D surface representation and SFM perception.

The outcome of all of these experiments is a novel appreciation of the role of area MT, which is commonly viewed as an area specialized for perceiving the direction of motion. We propose that it has a much more elaborate role, directly linked to 3-D SFM perception, and in particular with the segmentation and reconstruction of moving surfaces.

Transparency

Motion transparency is an interesting condition in which more than one surface can be represented at a single location in the visual field. Transparent stimuli have proven invaluable in exploring the surface reconstruction process and SFM perception. It has long been appreciated that motion thresholds are higher under transparent conditions[31]. In physiology experiments, we found that direction-selective V1 neurons generally give the same response to a stimulus moving in their preferred direction, whether or not a second stimulus is present and moving in the opposite direction[11]. On the other hand, we found strong suppression of MT neurons under these transparent conditions. This result led us to propose that the motion pathway contained at least two functional stages, one in V1 that measures motion signals, and a second, opponent stage in MT. It was proposed that this opponent stage was part of a surface reconstruction process, and functions to suppress noise such as motion flicker. This two-stage calculation resembles several current models of motion perception that use separate measurement and pooling/noise-reduction stages[32–35].

A prediction of this two-stage model is that completely 'balanced' stimuli, which contain motion signals in opposite directions at every local region in the display, should completely erase the perception of motion. We developed several displays which eliminated any motion perception. One of these consisted of two random dot patterns in which the dots moving in opposite directions were all paired with each other[36]. This stimulus produced an even greater suppression in area MT, providing strong evidence that area MT is the locus of perception of motion transparency[37]. Furthermore, the degree of suppression of individual MT neurons correlated negatively with their response to visual noise, supporting our proposal that one purpose of the opponency is to suppress noise during surface reconstruction.

Integration of disparity and motion

In our psychophysical studies with balanced, non-transparent stimuli, it was found that if the two surfaces were separated in depth, then the perception of the two surfaces reemerged[36]. Another demonstration of the strong interrelationship of stereo and motion is the study of Nawrot and Blake[38], which demonstrated that viewing SFM displays with stereo can subsequently bias the perceived depth order of ambiguous SFM stimuli.

Physiological experiments have established that area MT neurons are selective not only to motion, but also to stereoscopic depth[39]. We reasoned that the improved perception of transparency with stereoscopic depth separation may be a result of the opponent suppression in MT operating primarily within the same stereo planes. Our experiments showed that in fact this is true; inhibition in MT occurs mainly between motion signals with similar disparities[40]. This result is consistent with opponency operating to minimize random motion signals from a given surface (e.g. those due to flicker), while stereospecificity prevents opponency from occurring between motion signals from different objects at different visual depths. Figure 2 shows this basic MT network with opponent suppression within disparity channels, along with excitatory connections between opponent directions at different stereoscopic depths (Bradley, Qian and Andersen, unpublished observation). As we will see, this basic network may also be the neurological basis of the bistable percept that results for ambiguous stimuli like the rotating cylinders (Box 1 Fig.).

Neural correlates of bistable SFM perception

We trained monkeys to perform SFM tasks similar to those performed by our human subjects. After lesion to area MT these animals had a long lasting deficit in their ability to perceive SFM, even after simple motion thresholds recovered[41]. These were the first experiments showing that area MT plays a role in the SFM computation. Ideally, we wanted to have a more direct measure of the participation of MT in SFM perception. To accomplish this, we utilized the bistable aspect of the cylinder stimulus. Presenting the same physical stimulus can lead to two different perceptions. We reasoned that if MT activities varied consistently with the animals' perceptual reports, then this area has a direct role in SFM perception. We trained animals to view cylinder displays and then tell us which way they perceived the cylinders rotating by making eye movements in the perceived direction of motion of the front surface[42] (Fig. 3A). To be sure the animals did not cheat, on many randomly interleaved trials we added stereoscopic disparity to the dots. In these cases the cylinders are not ambiguous. Fig. 3B shows two animals' average performance on this task. When no disparity was added to the display, the animals' performance was at chance (50%). With even a small amount of disparity (12.5% of what would be expected from a real object) the animals performed well above chance, and with greater amounts of disparity the animals' performance continued to improve. This smooth and monotonically increasing psychometric function suggests that the animals are indeed reporting the direction of rotation they perceive.

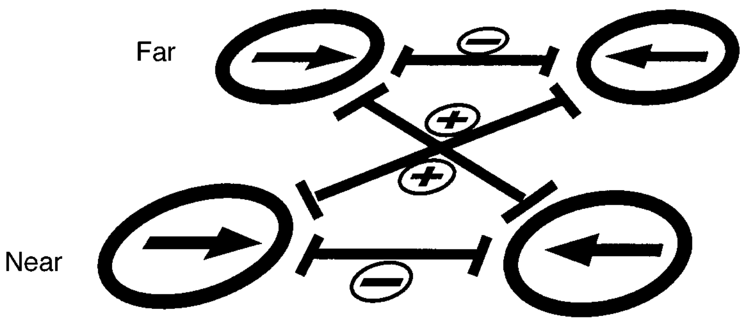

Fig. 2 Schematic circuit of direction- and disparity-dependent interactions in MT. Opposite motion directions suppress MT responses, but this suppression decreases, sometimes becoming facilitation, when the two directions appear in different disparity planes. This suggests inhibitory interactions between MT cells tuned for opposite directions and similar depths (horizontal lines) and excitatory connections between neurons tuned for opposite directions and different depths (diagonal lines).

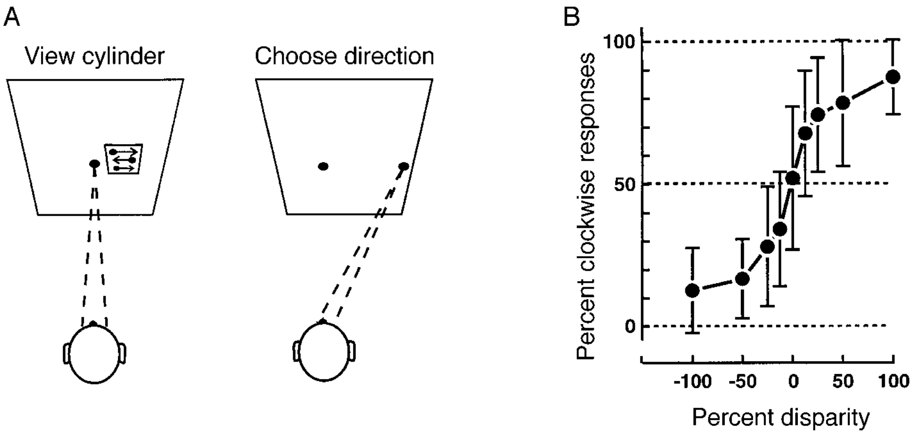

Fig. 3 Behavioral paradigm for studying neural mechanisms of structure-from-motion perception. (A) A monkey fixates a stationary spot while a cylinder projection, centered in the receptive field of the neuron under study, appears for 1 s. The monkey then looks at one of two targets to indicate the direction of the cylinder's *front* surface. **(B)** Mean performance of 2 animals (114 sessions). The ordinate shows the percentage of responses indicating 'clockwise' rotation (designated arbitrarily). The abscissa shows the amount of disparity in the cylinder as a percentage of the disparity of a real (3-D) cylinder; negative values denote 'counterclockwise' rotation (i.e. opposite to clockwise). As the percentage disparity moves away from zero, performance steadily improves, indicating that monkeys consistently tried to report the direction of rotation (and not, for example, resorting to guessing at low disparities). Error bars show standard deviation. Adapted from Bradley *et al.*[42]

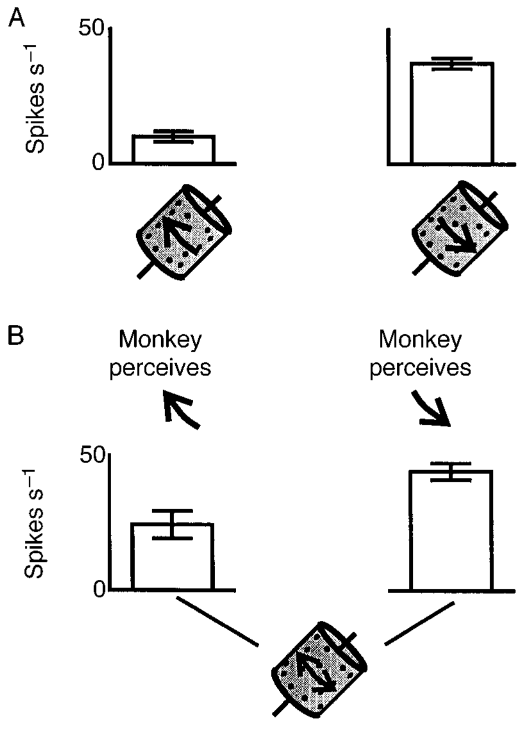

Fig. 4 Data from an MT neuron whose activity was correlated with the perception of structure-from-motion. (A) Responses to a full-disparity cylinder (unambiguous) rotating in opposite directions. This neuron clearly preferred the front moving down and to the right. **(B)** Responses to zero-disparity cylinders (bistable). Although the surface order (direction of rotation) was unspecified, responses were consistently higher when the monkey perceived the front surface moving down and to the right.

While the animals performed this psychophysical task, we recorded the activity of MT neurons. We found the preferred direction and disparity for each cell, then centered the cylinder stimulus in the receptive field, oriented such that one of the two surfaces was moving in the preferred direction of the cell. For instance the cell illustrated in Fig. 4 preferred down/right motion at near disparities. As predicted, this cell showed a larger response to full disparity cylinders in which the front surface moved down and to the right rather than up and to the left (Fig. 4A). For those trials in which there was no disparity added, when the monkey perceived the front surface moving down and to the right there was also more activity than when he perceived the front surface moving up and to the left (Fig. 4B). Thus, the cell's activity reflected, in a predictable way, the perception of the monkey in these bistable displays. Approximately one-third of MT cells showed this result, having a statistically significant difference in activity for the same stimulus, depending on the animal's perception. Moreover, 50% of the MT cells which responded differently to the full disparity cylinders showed this perceptual effect. Most of these 'perceptual' cells (27/34) showed the same variation in activity that was predicted from the responses to the full disparity cylinders. This correlation of perception with activity was true regardless of whether the cells responded best to their preferred direction in front (near cells; 17/20) or in back (far cells; 7/9; five cells could not be classified as near or far). Since such a small number of cells (7/68, 10%) showed significant activity differences that were 'anticorrelated' with perception, this may reflect the false-positive rate, which is expected to be around 5% (i.e. there may not be a distinct group of anticorrelated cells).

Figure 5A shows the mean normalized response, at each disparity, for the population of MT cells showing activity correlated with perception. This plot shows that the activity in the error trials at all disparities moved in the direction of the activity for the correct responses. Figure 5B shows the time course of the population activity for the 0% and 100% disparities. Note that the differential response for preferred versus non-preferred rotation for the bistable (zero-disparity) condition is similar to that for the full-disparity condition. Thus, the population response of these MT neurons is

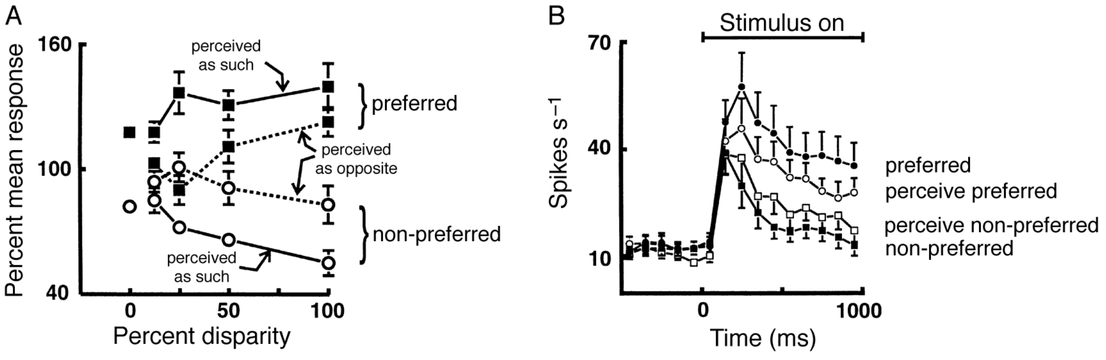

Fig. 5 Averaged data from 27 MT neurons whose activity was correlated with structure-from-motion perception. (A) Mean firing rates as a function of cylinder disparity. The top two curves (square symbols) represent trials where the neuron's preferred rotation was actually displayed; the bottom curves, where the non-preferred order was displayed. The outer curves (unbroken lines) represent correct trials; i.e. the answer matched the actual rotation. Inner traces (broken lines) are error trials. Whether the preferred or non-preferred order was shown, responses tended to be higher when the preferred order was *perceived*. Since zero-disparity cylinders do not have an 'actual' rotation, there are no correct and incorrect trials in this condition and the two data points at far left are plotted only as a function of the perceived rotation. **(B)** Time course of responses to full-disparity and zero-disparity cylinders. Circles represent the preferred direction of cylinder-rotation and the squares the non-preferred rotation. For the full-disparity cylinders (outer traces), responses diverge after 200–300 ms, with the higher trace corresponding to the neuron's preferred rotation (only data from correct trials shown). Responses to the zero-disparity cylinders show a similar time course, but in this case the higher trace corresponds to trials where the preferred order was *perceived*. Adapted from Bradley *et al.*[42]

strongly correlated with what the animal perceives in the bistable displays.

The circuitry for noise reduction during surface reconstruction, illustrated in Fig. 2, could be responsible for the bistable illusion. Most MT neurons prefer near or far stimuli, but their disparity tuning is broad enough to usually include zero disparity[39]. Thus, for example, a cylinder revolving around a vertical axis will activate four populations of MT neurons, tuned to near-right, near-left, far-left and far-right. However, because of the intrinsic excitatory and inhibitory circuitries discussed above, this situation is unstable and will tend toward a steady state that puts activity for opposite directions in different depth channels. For instance, if there is slightly more activity in the near-right cells, this would lead to suppression of near-left cells and activation of far-left cells that would in turn suppress far-right neurons – resulting in a perceived counter-clockwise rotation. Nawrot and Blake proposed a similar model based on psychophysical data[43].

These data show that MT responses directly reflect the perceived depth of moving surfaces and suggest that MT has a basic role in SFM perception. Of course, it is possible that SFM perception occurs in another area connected to MT, but even if this is the case, our results suggest that the perception of SFM is ultimately influenced by the segregation of MT activity into separate depth channels.

Conclusions

There has been considerable progress recently in understanding the neural algorithms responsible for SFM perception. Psychophysical experiments indicate that surfaces are reconstructed from motion measurements, and physiological experiments suggest that this computation is carried out in two stages. First, motion measurements are made in area V1. The next stage, surface reconstruction, occurs in MT, where direction opponency suppresses noise as part of the reconstruction process. However, the suppression mechanism is also 'smart' in not suppressing differential motion signals at different depths as indicated by disparity. Thus, disparity is used as a cue to determine when motions are from different surfaces and should not be suppressed. These observations, together, strongly support a role for MT in the segmentation and reconstruction of moving surfaces.

MT circuitry may also be responsible for the perception of depth induced by opposing retinal motions. Using 2-D projections of revolving cylinders, we found that MT responses are correlated with the perception of surface order, even when surface order is not specified in the stimulus. The mechanism of this perception may involve a segregation of activity in MT as a result of competing interactions which drive responses to opposing motions into separate depth channels. This is a new insight into the functions of an area that was previously thought to play a more rudimentary role in perceiving motion direction. SFM perception promises to be an exciting and fruitful area of research in the coming years.

Outstanding questions

- Are other cues besides disparity used by MT to segment moving surfaces? For instance, an interesting experiment would be to determine if luminance or color differences for transparent motions relieve opponent suppression in a manner similar to disparity.
- Are MT neurons sensitive to gradients of motion, and does MT activity correlate not only with the depth order of surfaces, but also with their perceived 3-D shape? Experiments by Treue and Andersen[44] examined the tuning of classical MT receptive fields for speed gradients and found that it was present, but modest. However, Xiao *et al.* showed strong tuning in MT for motion gradients in the surround, suggesting that the center-surround structure may support 3-D slant and curvature perception[45]. It has recently been shown that MT surrounds are opponent for disparity (unpublished data), providing additional support for surround involvement in coding depth gradients. Note, however, that curvature is given by the second spatial derivative of image velocity[46,47], so it will be important to determine whether MT neurons encode second derivatives of motion fields. Recent experiments by our group suggest that MT activity also correlates with perceived surface order for two flat sheets of dots moving over each other (unpublished data). It would be interesting to compare MT responses to cylinders and flat sheets on a cell-by-cell basis. If response differences (associated with surface order) are larger for cylinders, this would indicate that MT neurons have a role in coding surface shape. If not, it will be important to know whether shape is computed in MSTd, since many of these neurons are sensitive to motion patterns[26–30].
- There are also several interesting questions related to the motion-processing hierarchy. For instance, do V1 activities also correlate with perceived depth order in bistable displays? The relative lack of opponency in V1 (Refs 11,37), and the apparent importance of stereo-specific opponency for the bistable illusion, predict that they would not. But opponency does appear to occur in MSTd (Ref. 37), so activities in this area might well correlate with depth-order perception. MSTd neurons are sensitive to disparity[48], and to complex motion gradients[26–30], suggesting that this area could be involved in SFM perception. If perceptual effects are larger in MSTd than MT, such a result would be reminiscent of recent experiments by Logothetis and colleagues, who showed increasing perceptual correlation with activity in the ventral pathway for rivalrous stimuli when proceeding from areas V1/V2 to V4 to IT (Refs 49,50).
- Are similar mechanisms and cortical areas involved in human and monkey SFM perception. fMRI experiments in humans will be useful in examining this question.

References

1 Miles, W.R. (1931) Movement interpretations of the silhouette of a revolving fan *Am. J. Psychol.* 43, 392–405

2 Wallach, H. and O'Connell, D.N. (1953) The kinetic depth effect *J. Exp. Psychol.* 45, 205–217

3 Clocksin, W.F. (1980) Perception of surface slant and edge labels from optical flow: a computational approach *Perception* 9, 253–269

4 Longuet-Higgins, H.C. and Prazdny, K. (1980) The interpretation of a retinal moving image *Proc. R. Soc. London Ser. B* 208, 385–397

5 Koenderink, J.J. and Van Doorn, A.J. (1986) Depth and shape from differential perspective in the presence of bending deformations *J. Opt. Soc. Am. Ser. A* 3, 242–249

6 Husain, M., Treue, S. and Andersen, R.A. (1989) Surface interpolation in three-dimensional structure-from-motion perception *Neural Comput.* 1, 324–333

7 Hildreth, E.C. *et al.* (1995) Recovering three-dimensional structure from motion with surface reconstruction *Vis. Res.* 35, 117–137

8 Ullman, S. (1984) Maximizing rigidity: the incremental recovery of a 3-D structure from rigid and non-rigid motion *Perception* 13, 255–274

9 Grzywacz, N.M. and Hildreth, E.C. (1987) Incremental rigidity scheme for recovering structure from motion: position-based versus velocity-based formulations *J. Opt. Soc. Am. Ser. A* 4, 503–518

10 Grzywacz, N.M. *et al.* (1988) The temporal integration of 3-D structure from motion: a computational and psychophysical study, in *Organization of Neural Networks* (VonSeelen, W., Shaw, G. and Leinhos, U.M., eds), pp. 239–259, Weinheim

11 Snowden, R.J. *et al.* (1991) The response of area MT and V1 neurons to transparent motion *J. Neurosci.* 11, 2768–2785

12 Shariat, H. and Price, K.E. (1990) Motion estimation with more than two frames *IEEE Patt. A* 12, 417–434

13 Dosher, B.A., Landy, M.S. and Sperling, G. (1989) The kinetic depth effect and optic flow: 1. 3-D shape from Fourier motion *Vis. Res.* 29, 1789–1813

14 Landy, M.S. *et al*. (1991) The kinetic depth effect and optic flow: 2. First-order and second-order motion *Vis. Res*. 31, 859–876

15 Treue, S., Husain, M. and Andersen, R.A. (1991) Human perception of structure from motion *Vis. Res.* 31, 59–75

16 Mckee, S.P. and Welch, L. (1985) Sequential recruitment in the discrimination of velocity *J. Opt. Soc. Am. Ser. A* 2, 243–251

17 Nakayama, K. (1985) Biological image motion processing: a review *Vis. Res.* 25, 625–660

18 Andersen, R.A. and Siegel, R.M. (1990) Motion processing in primate cortex, in *Signal and Sense: Local and Global Order in Perceptual Maps* (Edelman, G.M., Gall, W.E. and Cowan, W.M., eds), pp. 163–184, John Wiley & Sons

19 Maunsell, J.H. and Van Essen, D.C. (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey: I. Selectivity for stimulus direction, orientation and speed *J. Neurophysiol*. 49, 1127–1147

20 Newsome, W.T. and Wurtz, R.H. (1988) Probing visual cortical function with discrete chemical lesions *Trends Neurosci*. 11, 394–400

21 Movshon, J.A. *et al*. (1985) The analysis of moving visual patterns, in *Study Group on Pattern Recognition Mechanisms* (Chagas, C., Gattass, R. and Gross, C.G., eds), pp. 117–151, Pontifica Academia Scientiarium, Vatican City

22 Lappin, J.S. Doner, R. and Kottas, B.L. (1980) Minimal conditions for the visual detection of structure and motion in three dimensions *Science* 209, 717–719

23 Hubel, D.H. and Wiesel, T.N. (1968) Receptive fields and functional architecture of monkey striate cortex *J. Physiol*, 195, 215–243

24 Allman, J.M. and Kaas, J.H. (1971) A representation of the visual field in the caudal third of the middle temporal gyrus of the owl monkey (*aotus trivirgatus*) *Brain Res.* 31, 85–105

25 Zeki, S.M. (1974) Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey *J. Physiol.* 236, 549–573

26 Saito, H. *et al*. (1986) Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey *J. Neurosci*. 6, 145–157

27 Sakata, H., Shibutani, H. and Kawano, K. (1983) Functional properties of visual tracking neurons in posterior parietal association cortex of the monkey *J. Neurophysiol*. 49, 1364–1380

28 Duffy, C.J. and Wurtz, R.H. (1991) Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli *J. Neurophysiol*. 65, 1329–1345

29 Graziano, M.S., Andersen, R.A. and Snowden, R.J. (1994) Tuning of MST neurons to spiral motions *J. Neurosci*. 14, 54–67

30 Lagae, L. *et al*. (1994) Responses of macaque STS neurons to optic flow components: a comparison of areas MT and MST *J. Neurophysiol*. 71, 1597–1626

31 Mather, G. and Moulden, B. (1983) Thresholds for movement direction: two directions are less detectable than one *Q. J. Exp. Psychol*. 35, 513–518

32 Watson, A.B. and Ahumada, A.J. (1985) Model of human visual-motion sensing *J. Opt. Soc. Am. A* 2, 322–342

33 Vansanten, J.P.H. and Sperling, G. (1985) Elaborated Reichardt detectors *J. Opt. Soc. Am. A* 2, 300–321

34 Adelson, E.H. and Bergen, J.R. (1985) Spatiotemporal energy models for the perception of motion *J. Opt. Soc. Am. A* 2, 284–299

35 Qian, N., Andersen, R.A. and Adelson, E.H. (1994) Transparent motion perception as detection of unbalanced motion signals: 3. Modeling *J. Neurophysiol*. 14, 7381–7392

36 Qian, N. and Andersen, R.A. (1994) Transparent motion perception as detection of unbalanced motion signals: 1. Psychophysics *J. Neurophysiol*. 14, 7357–7366

37 Qian, N. and Andersen, R.A. (1994) Transparent motion perception as detection of unbalanced motion signals: 2. Physiology *J. Neurophysiol*. 14, 7367–7380

38 Nawrot, M. and Blake, R. (1991) The interplay between stereopsis and structure from motion *Percept. Psychophys*. 49, 230–244

39 Maunsell, J.H. and Van Essen, D.C. (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey: II. Binocular interactions and sensitivity to binocular disparity *J. Neurophysiol*. 49, 1148–1167

40 Bradley, D.C., Qian, N. and Andersen, R.A. (1995) Integration of motion and stereopsis in middle temporal cortical area of macaques *Nature* 373, 609–611

41 Siegel, R.M. and Andersen, R.A. (1986) Motion perceptual deficits following ibotenic acid lesions of the middle temporal in the behaving rhesus monkey *Soc. Neurosci. Abstr*. 12, 1183

42 Bradley, D.C., Chang, G. and Andersen, R.A. (1998) Encoding of 3-D structure-from-motion by primate area MT neurons *Nature* 392, 714–717

43 Nawrot, M. and Blake, R. (1991) A neural-network model of kinetic depth *Visual Neurosci*. 6, 219–227

44 Treue, S. and Andersen, R.A. (1996) Neural responses to velocity-gradients in macaque cortical area MT *Visual Neurosci*. 13, 797–804

45 Xiao, D.K. *et al*. (1997) Selectivity of macaque MT/V5 neurons for surface orientation in-depth specified by motion *Eur. J. Neurosci*. 9, 956–964

46 Leone, G., Droulez, J. and Cornilleau-Peres, V. (1992) A neural network for computing surface curvature from optic flow *Ann. New York Acad. Sci.* 656, 861–864

47 Cornilleau-Peres, V. and Droulez, J. (1993) Stereo–motion cooperation and the use of motion disparity in the visual perception of 3-D structure *Percept. Psychophys*. 54, 223–239

48 Roy, J.P. and Wurtz, R.H. (1990) The role of disparity-sensitive cortical neurons in signaling the direction of self-motion *Nature* 348, 160–162

49 Leopold, D.A. and Logothetis, N.K. (1996) Activity changes in early visual cortex reflect monkeys percepts during binocular-rivalry *Nature* 379, 549–553

50 Shenberg, D.L. and Logothetis, N.K. (1997) The role of temporal cortical areas in perceptual organization *Proc. Natl. Acad. Sci. U. S. A.* 94, 3408–3413