INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

SMR.853 - 73

ANTONIO BORSELLINO COLLEGE ON NEUROPHYSICS

(15 May - 9 June 1995)

"Visual processing of motion"

David C. Burr
Istituto di Neurofisiologia
Consiglio Nazionale delle Ricerche
56100 Pisa
Italy

These are preliminary lecture notes, intended only for distribution to participants.

Reprinted from Trends in Neuro Sciences – *July, 1986* Vol. 9, No. 7

Visual processing of motion

Image motion does not pose the same problem for the eye as it does for the camera. We are able to see objects with clarity when they move. We can also see motion on cinema or TV when what we are shown is a sequence of stills. As followers of fast ball-games like tennis and cricket know, we can see motion sequences as sharp when the actual individual stills on the cinema or TV screen, viewed singly or at slow speed, contain blurred images of moving actors or objects. Only recently has the idea emerged that visual abilities such as these may be explained by the specializations of visual neurones. Theoretical and experimental advances in neurophysiology are beginning to identify the tuning characteristics, in space and time, of mechanisms that make it possible to do what a camera cannot: resolve form and motion simultaneously.

The visual system responds not to instantaneous distributions of light intensity imaged on the retinae, but to integrals of them over a period of about ⅒ s (Refs 1–3). One might suppose from this that all but the most slowly moving objects should always appear blurred, as in a still photograph taken with a ⅒ shutter speed (see Fig. 1): but they do not. Blur has been measured in several ways, most tellingly as a function of exposure time[4]. The length of the visual streak left behind by moving spots increases with exposure time up to 30 ms. Beyond that the length decreases, and at exposures of ⅒ s and longer, spots moving at speeds up to 16 deg s^{-1} leave no streak at all. They appear as spots in motion, with no comet-like tail. Thus the visual system is capable of resolving moving images as they really are, without smear, even when the eye does not track them, but to do so it requires an exposure time of 100 ms or more. This suggests that accurate perception of form in motion involves visual integration of a type which eliminates smear.

The manner in which vision summates energy from moving objects reveals more about visual integration[3]. When spots move, the minimum light energy required for visibility decreases with exposure up to ⅒ s, during which period the spot traverses 1.6 deg (100 ms at 16 deg s^{-1}), a distance comfortably exceeding the range of summation for stationary discs[1] or lines[3]. Thus motion invokes mechanisms that collect light from moving targets and that do so over larger stretches of space than mechanisms dealing with stationary targets. As Barlow[5,6] has pointed out, these experiments (and those on motion smear) imply that the visual system integrates along a trajectory in space and time matching that of the moving object, not separately in time and in space, as might have been supposed. This in turn suggests a variety of mechanisms tuned to different image sizes and different image speeds.

It has long been known that the cinematograph, which produces a succession of discrete stills, simulates motion. Early in the century it was suggested that special visual processes were required to explain the apparent motion of cinema[7]. The issue has remained a lively one. Recent experiments[8–12] show not only that cinematic motion appears as real, but also that the visual system interpolates along the path of apparent motion to perceive objects occupying positions intermediate to those displayed in the discrete frames. Furthermore, whole forms can be synthesized by interpolation. This capacity to synthesize forms the basis of a visual display device called the 'betagraph'[13] which is illustrated in Fig. 2. By the method explained in the caption to Fig. 2, the message strip of part A may be viewed through the slits of part B of the figure. The word 'Vision' is clearly visible, although the narrow slits ensure that the whole of the word is never simultaneously all in view. The vernier is also seen, although only one half is ever displayed at a time. The hidden information is interpolated by the motion mechanisms of vision, and interpolated with precision. Stereo[8,9] and vernier acuities[10–12], the finest known, are almost as good under these conditions as when the two components are presented simultaneously.

What are the mechanisms of vision which allow for integration along a spatio-temporal trajectory when there is real motion, and interpolation when motion is sampled, as in cinema? Visual neurones in cat and monkey cortex are selectively tuned for spatial and temporal frequency, as well as for the orientation of visual stimuli. That is, they respond best to oriented patterns of appropriate spatial frequency (size) and temporal frequency (rate) (see for example Refs 15–17). Although one cannot measure directly the selectivity of individual neurones in human cortex, a variety of indirect techniques demonstrate that human visual neurones are similarly tuned[18–21]. Once the properties of such visual units are understood it is possible to see how they integrate to give information about form and motion, and how they interpolate.

Using one indirect technique, masking, we have recently determined the spatio-temporal tuning function of human motion detectors[22]. We measured the minimum contrast necessary for an observer to detect the direction of

Fig. 1. *A photograph taken with 125 ms shutter speed, about the same as the summation period of the visual system[1–3]. Visual scenes do not normally appear so blurred. This implies mechanisms that can analyse form of objects while they are in motion[4].*

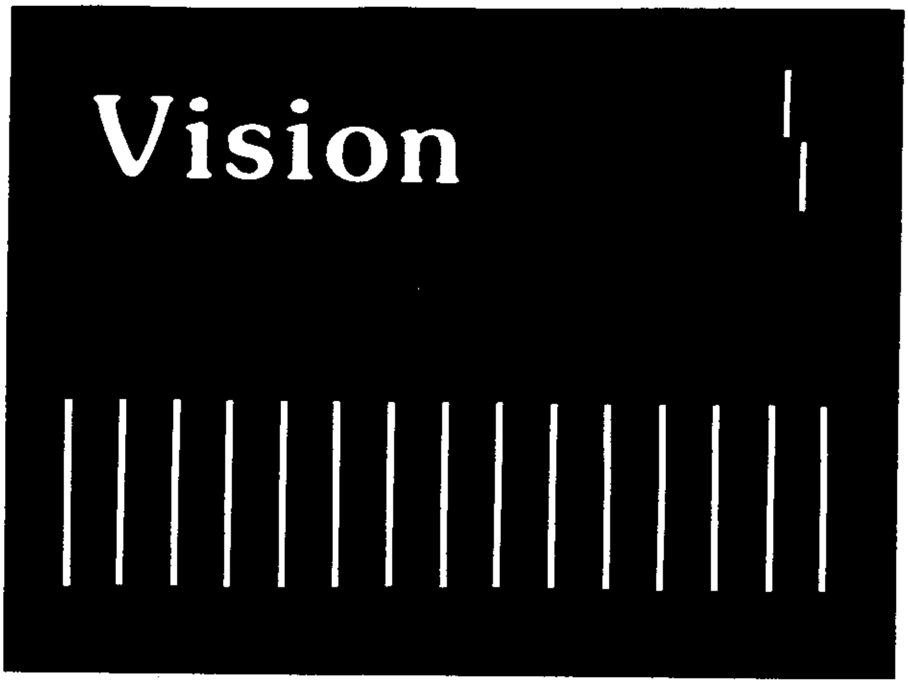

Fig. 2. *Interpolation by motion mechanisms of objects shown only in fixed positions to intermediate positions along the path of (apparent) motion is the basis of a display device called the 'Betagraph'[13], illustrated here. It differs from conventional moving signs in that the whole of the message is never displayed. Only the fragments showing through narrow slits are visible; yet the message appears whole because of visual interpolation. Photocopy the figure onto a transparency, separate the two parts and superimpose them on an overhead projector (or simply hold them to the light). Hold part B still and move part A horizontally across it. The word VISION is clearly visible, though the slits ensure that all the different parts of it are never simultaneously in view. The vernier is also seen, with an offset between top and bottom lines, though the two are never simultaneously displayed. Interpolation is done with precision. Stereo and vernier acuities, the finest known, are almost as good under these conditions as in normal stationary presentation[8–12]. Eye movements aid the process of interpolation in this crude demonstration[14]; but controlled experiments show that they are neither necessary nor sufficient for high-precision interpolation[10]. Display systems based on the principle illustrated here lend themselves to electronic embodiment, using LEDs arranged in strips.*

drift of a vertical sinusoidal grating of given spatial frequency and drift speed. We then remeasured sensitivity to the same grating while 'mask' gratings of varying spatial and temporal frequency were superimposed. The idea is that if a mask grating is sufficiently similar to the test in spatial and temporal frequency to stimulate the same neurones that detect the test, it will lower sensitivity to the test[20]. Masks can thus be used to probe the sensitivity of mechanisms stimulated by the test. Appropriately scaled, the suppression of sensitivity caused by the masks describes the spatio-temporal frequency selectivity profiles of human motion detectors. The selectivity profile of one representive motion detector is depicted in the contour plot of Fig. 3A. The function is peaked at a spatial frequency near 1 c deg^{-1} and a temporal frequency near 8 Hz, the frequencies of the test grating. Other detectors peak at different spatial frequencies; but all peak near 10 Hz.

While Fig. 3A describes fully the response of one hypothetical motion detector to variations in the spatial and temporal frequency of stimuli, it does little to aid our intuitions, which are more attuned to the familiar domain of space and time. At the cost of assuming linearity, we can convert the frequency tuning function into a space–time function by a simple mathematical manipulation, the Fourier transform. This gives us the response profile shown schematically in Fig. 3B. A conventional receptive field is a response profile in two dimensions of space. It describes the distribution of light in space to which a cell is tuned. It is no great step to imagine a field that requires a distribution of light to move for maximal stimulation. The response profile shown in Fig. 3B may be considered, by an extension of the conventional concept, as the spatio-temporal receptive field of a motion detector (see also Refs 23–26). For simplicity we consider only one spatial dimension. The field is oriented in space–time as the classic 2-dimensional spatial receptive field is oriented in space–space[27]. The orientation gives it a velocity tuning (see figure caption). The field of this particular detector, tuned to 1 c deg^{-1} and 8 Hz, is oriented along a velocity axis of 8 deg s^{-1}. Others will be of different frequency tuning and oriented at different angles, reflecting tuning for different sizes and velocities[22] (including those in the opposite direction).

Mechanisms of this sort integrate along the trajectory of motion, providing for simultaneous analysis of form and motion. An object moving at the velocity corresponding to the orientation of the receptive field will not leave a streak of activity on 'the mind's eye'. Rather, as it moves, it will stimulate to maximum response mechanisms which, by their design, signal form in motion. The longer an image traverses the receptive field, the stronger and more numerous will be the signals carrying information about its content and how it is moving. Such a scheme readily explains why we see movement without blur despite a long summation time. The extent of the field is the stretch of space over which it collects energy from a moving image. Its resolution of the moving image is determined by the internal structure of the field, along the direction of motion. Further precision of resolution can be achieved by the co-operative action of many neurones with fields of different cross-sectional profile.

The concept of receptive fields oriented in space–time accommodates cinematic and other forms of apparent motion. Apparent motion has often been considered as motion the brain must derive by computational processes from successive stills. And, indeed, real motion has sometimes been treated as if it, too, posed a computational problem[28–30]. Cinematic, stroboscopic and other forms of apparent motion may be considered as forms of sampled motion and be represented (Fig. 3B) as a series of dots or dashes in the space–time domain. As far as motion detectors can tell, this motion is true motion. The detectors cannot distinguish discrete from smooth; they simply signal the stimulation they receive. Recent experiments show that motion will appear smooth unless sampling is sufficiently coarse to generate sampling artefacts at frequencies which can be detected by other mechanisms[31].

The tuning characteristics of motion detectors make simulated motion possible; they also restrict the range of motion we can detect. We can see the second hand of our watch move, but not the minute hand. A disc of high spatial frequency grating at low contrast will

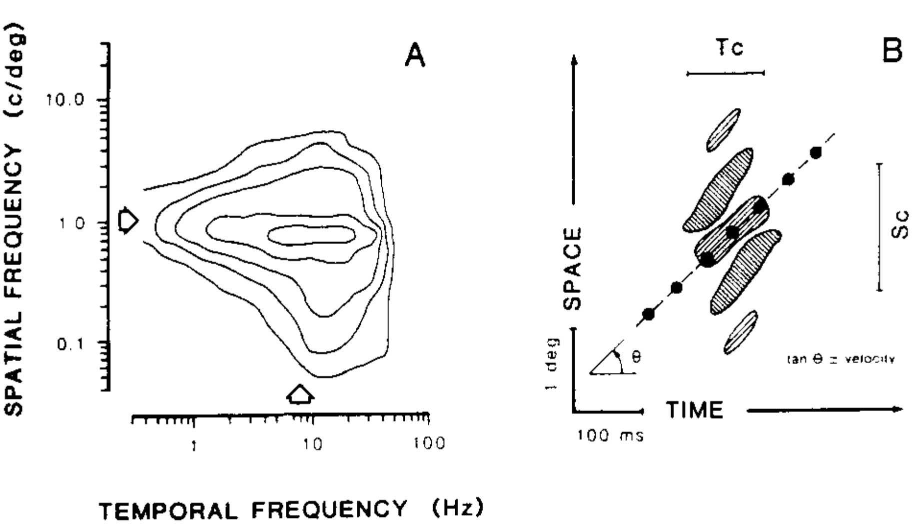

Fig. 3. **(A)** *The spatio-temporal tuning function of one class of motion detector, obtained by the psychophysical technique of masking*[22]*. Observers adjusted the contrast of a drifting vertical test grating until its direction of drift could just be discerned, giving a measure of contrast sensitivity. The grating had a spatial frequency of 1 c* deg^{-1} *and drifted at 8 Hz (indicated by the arrows). Contrast sensitivity was then measured again in the presence of superimposed mask gratings of various spatial frequencies, reversing in contrast at various temporal frequencies. Masks similar to the test in both spatial and temporal frequency raise test threshold, implying that they stimulate the same neurones that detect the test, lowering their sensitivity (or gain)*[20]*. Masks differing greatly in spatial or temporal frequency do not affect test threshold, implying that they are outside the selectivity bandwidth of the neurones which detect the test. After appropriate scaling, the degree of masking at each spatial and temporal mask frequency gives the spatio-temporal selectivity function of the class of neurones stimulated by the test grating. This is depicted by the contour map of* **(A)**. *The function is peaked at 1 c* deg^{-1}*, 8 Hz, and falls off steadily for higher or lower spatial or temporal frequencies (each contour line represents 0.5 log units). The tuning function of* **(A)** *gives the frequency response of a motion neurone, but does little to aid our intuition as to the types of moving images to which it will respond. To visualize the spatio-temporal structure of the receptive field of the detector, it is necessary to derive its impulse response function, by Fourier transform of the tuning function* **(B)**. *The necessary assumptions are system linearity and linear phase. Other assumptions about phase (e.g. minimum phase) yield very similar pictures of receptive fields*[22]*. We refer to the impulse response function as the spatio-temporal receptive field of a motion detector, following the conventions of neurophysiology. Like the receptive fields of simple cells, the field is made up of several adjacent opponent regions (indicated by orthogonal hatching). The field may be either centre ON (preferring light at its centre) or centre OFF (preferring dark at its centre), depending on the phase assumptions. However, whereas typical representations of simple cells are in space–space, this field lies in the space–time domain. Orientation in space–time does not refer to physical orientation, but to velocity (given by the tangent of the angle from the time axis). The long axis of this particular field is oriented at velocity 8 deg* s^{-1} *(indicated by the dotted line), the velocity of the test grating. Fields of other detectors will be oriented more or less steeply, reflecting different velocity tuning (which ranges from 1 to about 1000 deg* s^{-1}*, Ref. 22). TC refers to the extent of summation in time (defined as the extent of the receptive field in the time dimension up to half height) and Sc to the extent of summation in space. The estimates shown in the figure are reasonably consistent with measurements of summation*[3]*. Spatial summation of this field does not imply smear, as the summation occurs along a velocity trajectory, not separately in the two domains. The dots represent sampled motion at 24 Hz, the rate of commercial cinema. The detector will summate the samples of motion along the motion trajectory to signal continuous motion.*

not appear to rotate at all[32] (stopped motion) since stimulation is outside the range of any motion detector. There is evidence[33] that during saccades, the ballistic eye-movements we make to shift gaze, the tuning range of motion detectors is shrunk, leaving us blind to the rapid image motion that saccades produce, and letting the visual world appear stable. Saccades blunt motion sensitivity without impairing other visual sensitivities, suggesting functional independence of motion detectors. This idea is strengthened by clinical reports of a patient whose vision is normal in all other respects but who cannot see moving objects, even people entering a room[34]. The patient is also blind to apparent motion, both 'short range' and 'long range'[35] as would be expected if the same mechanisms underlie perception of real and apparent motion.

This new concept of motion perception, derived from psychophysical measurements in man, receives support from several recent neurophysiological studies of cat cortical neurones. Emerson and his colleagues have recorded the response of complex cells in cat primary visual cortex to successively presented bars, and shown their spatio-temporal receptive fields have virtually the same form as that shown in Fig. 3B[36]. Further evidence suggests that the structure of the spatio-temporal receptive fields and their associated directional selectivity arise from local inhibitory interactions[37,38], along the lines suggested by Barlow and Levick[39] some twenty years ago. Furthermore, Swindle and Cynader (pers. comm.) have evidence that individual neurones are capable of spatio-temporal interpolation, responding to vernier offset created by temporal offset (like that of Fig. 2).

In human vision there must exist thousands of individual motion detectors, their receptive fields so structured that each detector will respond best to stimuli in a given region of space, and of given spatial frequency (size), speed and direction of motion. Our evidence indicates that in the temporal domain, motion detectors are all fairly similar: they are all tuned to a temporal frequency of about 10 Hz, and summate for about 100 ms[3,21,22]. In the spatial domain, however, there is great variation both in preferred spatial frequency and in size of receptive fields. Preferred spatial frequencies range from 0.025 to 15 c deg^{-1} (Ref. 21), and receptive field size from 2 s arc to 7 deg (Ref. 40) (larger fields preferring lower spatial frequencies). One implication of this is that detectors will vary considerably in their preference for stimulus speed allowing us to see clouds drift and snails crawl, as well as, under extreme experimental conditions, motion at speeds of over 10000 deg s^{-1} (Ref. 41). The existence of detectors with large receptive fields may help explain 'long range apparent motion'[35], which has been thought to involve processes separate from those for normal motion and apparent motion.

It is now well understood how the collective activity of differently tuned neurones explains visual sensitivity. It is beginning to be understood how these same neurones derive information from visual images. Extending the concept of the receptive field of a neurone into the space–time domain removes much of the mystery from motion by bringing it within a conceptual framework which has proved so fruitful for visual science.

Selected references

1 Barlow, H. B. (1958) *J. Physiol. (London)* 141, 337–350
2 Ross, J. and Hogben, J. H. (1974) *Vision Res.* 14, 1195–1201
3 Burr, D. C. (1981) *Proc. R. Soc. London, Ser. B* 211, 321–339
4 Burr, D. C. (1980) *Nature (London)* 284, 164–165

5 Barlow, H. B. (1979) *Nature (London)* 279, 189–190

6 Barlow, H. B. (1981) *Proc. R. Soc. London, Ser. B* 212, 1–34

7 Wertheimer, M. (1912) *Z. Psychol.* 61, 151–265

8 Burr, D. C. and Ross, J. (1979) *Vision Res.* 19, 479–484

9 Morgan, M. J. and Thompson, P. (1975) *Perception* 4, 3–18

10 Burr, D. C. (1979) *Vision Res.* 19, 835–837

11 Fahle, M. and Poggio, T. (1981) *Proc. R. Soc. London, Ser. B* 213, 451–477

12 Morgan, M. J. and Watt, R. J. (1983) *Vision Res.* 23, 997–1004

13 Ross, J. (1977) *J. Physiol. (London)* 271, 2–3

14 Morgan, M. J. (1981) in *Eye Movements: Cognition and Visual Perception* (Fisher, D. F., Monty, R. A. and Senders, J. W., eds), pp. 111–333, Erlbaum

15 Maffei, L. and Fiorentini, A. (1973) *Vision Res.* 13, 1255–1267

16 Tolhurst, D. J. and Movshon, J. A. (1975) *Nature* 257, 674–675

17 Movshon, J. A., Thompson, I. D. and Tolhurst, D. J. (1978) *J. Physiol. (London)* 283, 53–99

18 Blakemore, C. and Campbell, F. W. (1969) *J. Physiol. (London)* 225, 437–455

19 Sachs, M. B., Nachmias, J. and Robson, J. G. (1971) *J. Opt. Soc. Am.* 61, 1176–1183

20 Legge, G. E. and Foley, J. M. (1980) *J. Opt. Soc. Am.* 70, 1458–1471

21 Anderson, S. J. and Burr, D. C. (1985) *Vision Res.* 25, 1147–1154

22 Burr, D. C., Ross, J. and Morrone, M. C. (1986) *Proc. R. Soc. London, Ser. B* 227, 249–265

23 Burr, D. C. (1983) *Proc. IUPS XV*, 510.04

24 Watson, A. B. and Ahumada, A. J. (1985) *J. Opt. Soc. Am.* A2, 322–341

25 Adelson, E. H. and Bergen, J. R. (1985) *J. Opt. Soc. Am.* A2, 284–299

26 van Santen, J. P. H. and Sperling, G. (1985) *J. Opt. Soc. Am.* A2, 300–321

27 Hubel, D. H. and Wiessel, T. N. (1962) *J. Physiol. (London)* 160, 106–154

28 Ullman, S. (1983) *Trends NeuroSci.* 6, 177–179

29 Ullman, S. (1979) *The Interpretation of Visual Motion*, MIT Press

30 Anstis, S. M. (1980) *Philos. Trans. R. Soc. London, Ser. B* 290, 153–168

31 Burr, D. C., Ross, J. and Morrone, M. C. (1986) *Vision Res.* 26, 643–652

32 Campbell, F. W. and Maffei, L. *Nature* 278, 55

33 Burr, D. C., Holt, J., Johnstone, J. R. and Ross, J. (1982) *J. Physiol. (London)* 333, 1–15

34 Zihl, J., von Cramon, D. and Mai, N. (1983) *Brain* 106, 313–340

35 Braddick, O. (1980) *Philos. Trans. R. Soc. London, Ser. B* 290, 137–151

36 Emerson, R. C., Citron, M. C. and Vaughn, W. J. (1985) *Invest. Ophthalmol. Visual Sci. (Supp.)* 26, 7

37 Emerson, R. C., Citron, M. C. and Felleman, D. J. (1985) in *Models of the Visual Cortex* (Rose, D. and Dobson, V. G., eds), pp. 240–431, John Wiley & Sons

38 Swindale, N. V. and Cynader, M. S. (1984) *Soc. Neurosci. Abstr.* 10, 471

39 Barlow, H. B. and Levick, W. R. (1965) *J. Physiol. (London)* 178, 447–504

40 Anderson, S. J. and Burr, D. C. (1985) *Perception* 14, A19

41 Burr, D. C. and Ross, J. (1982) *Vision Res.* 23, 3567–3569

DAVID BURR AND JOHN ROSS

Department of Psychology, University of Western Australia, Perth WA 6009, Australia.

4.