
Temporal Summation of Moving Images by the Human Visual System

D. C. Burr

Proc. R. Soc. Lond. B 1981 **211**, 321-339

doi: 10.1098/rspb.1981.0010

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

To subscribe to *Proc. R. Soc. Lond. B* go to:
<http://rspb.royalsocietypublishing.org/subscriptions>

Temporal summation of moving images by the human visual system

BY D. C. BURR†

*Craik Laboratory, Department of Experimental Psychology,
University of Cambridge, Cambridge CB1 3DS, U.K.*

(Communicated by H. B. Barlow, F.R.S. – Received 8 July 1980)

Measurements of threshold visibility were made as a function of duration of stimulus exposure for small moving dot targets, drifting sinusoidal gratings and moving patches of sinusoidal gratings, to investigate how the human visual nervous system summates over time signals arising from stimuli in motion.

At image speeds of less than 16 deg/s, temporal summation is as strong and as extended for moving as for stationary dots (total summation over to about 100 ms). This summation is about twice that which would be expected from separate consideration of the regions of spatial and temporal integration.

Measurements with sinusoidal gratings reveal that the nature of the summation depends critically on the spatial frequency of the stimulus: gratings of low spatial frequency summate well when in motion (and *only* when in motion), whereas those of high spatial frequency summate well only when stationary or in very slow motion. An analogue simulation with electronic filters showed that these psychophysical results are directly predictable from the known transfer characteristics of the human visual system (with the additional assumption of probability summation at threshold).

Finally, with small patches of sinusoidal grating, it was established that translation *per se* across the retina has little effect on temporal summation. This suggests that the results obtained with sinusoidal gratings of large extent are also relevant to small moving stimuli, allowing the summation results obtained with dot stimuli to be discussed in terms of the temporal transfer properties of spatially selective visual detectors. On the basis of these results it is proposed that the extended temporal summation observed for dots in motion results from summation of energy of low spatial frequency present in these stimuli.

INTRODUCTION

It is well known that the visual system summates signals over time, presumably to obtain reasonable signal-to-noise levels. This was first demonstrated by Bloch (1885), who showed that the visibility of a brief flash of light depends only on the

† The author currently holds a Royal Society European Exchange Fellowship at the Laboratorio di Neurofisiologia del C.N.R., Via S. Zeno, 51–56100 – Pisa, Italy.

total energy of the flash, independently of its duration. Later, Graham & Margaria (1935), Barlow (1958) and others showed that the reciprocity between time and luminance holds only for up to some *critical duration*, whereafter visibility is determined solely by the luminance, rather than the energy, of the flash. The critical duration varies considerably with both the luminance and size of the stimulus, being longest for small stimuli, under dim background luminance (Barlow 1958).

Classically it was supposed that the summation results from *integration* of light signals over a fixed period of time corresponding to the critical duration. However, although integration may account for the time–luminance reciprocity of single flashes of light, it fails to explain some of the more complex interactions, such as those observed by Ikeda (1963, 1965) with double flashes of light. At certain inter-flash intervals (about 50 ms, depending on luminance) two successive flashes do not summate positively to lower the visibility threshold, but rather inhibit each other, resulting in a combined threshold, which is higher than either separate threshold. Furthermore, at this inter-flash separation, a flash of light (a brief increment of light) summates positively with a brief *decrement* of light. More recently, Watson & Nachmias (1977) have similarly shown that, when separated in time by about 50 ms, sinusoidal gratings summate best when in counterphase. Clearly, neither of these results can be explained solely by visual integration, as the integral with respect to time, both of an increment and decrement of light and of two gratings in counterphase, is zero.

A more general concept, which accounts for both summation and inhibition at various temporal separations, is *linear filtering* (De Lange 1954; Sperling & Sondhi 1968; Kelly 1971). A filter is a device that, by both integration and differentiation, passes only a limited band of frequencies. It may then be followed by some nonlinear operation, such as integration of the square of its output (Rashbass 1970) or probability summation (Tolhurst 1975; Watson 1979).

Almost all investigations to date have measured temporal or spatial summation only for stationary targets. However, for most visual animals, objects of paramount survival importance are often in motion, and if these are to be clearly resolved, especially under low contrast (camouflage) conditions, their images must also be summated over time. But here the visual system is faced with seemingly conflicting demands: on the one hand, efficient resolution of the *object* in motion requires pooling or averaging of signals over time; on the other, efficient detection of the *motion* requires a fast, transient response. Put another way, integration necessarily implies loss of temporal information, but this information is essential for motion perception.

To discover how this apparent contrariness is resolved, I have made some measurements of the visual summation of moving targets. These measurements reveal that the visual system has in fact succeeded in optimizing both factors simultaneously, and indeed it turns out that there is no conflict of interests between sensitivity to motion and resolution of structure. It seems that there exist visual

Temporal summation of moving images

323

mechanisms designed both to respond to motion and to pool over time signals from the moving target, thereby rendering it more visible. Evidence is presented to suggest that these mechanisms operate not solely by integration (which would encumber motion sensitivity) but also by a process better described as resonance of temporally tuned filters.

MOVING SPOTS OF LIGHT

First, I measured the visibility thresholds for a small moving light source, as a function of the exposure duration. The general procedure was like those standardly used in classical summation studies (see, for example, Barlow 1958), except that the stimulus was caused to move horizontally at a constant velocity. The measurements were made under photopic conditions, 7° above the fovea (where the retina is presumably more homogeneous (see, for example, Robson & Graham 1980)).

Methods

Observers sat 2.4 m from a 1 m^2 screen, illuminated by front projection to 30 cd/m^2 . The stimulus was generated on a small laboratory PDP-8/I computer and displayed on the face of a Textronix 602 point-plotting oscilloscope (equipped with fast-fade p15 phosphor). At the observer's initiation, the stimulus (a small spot of light) moved stroboscopically either to the left or the right (the direction being randomized between trials) at a speed ranging from 0 to 32 deg/s ; that is, it was displaced laterally a certain distance every 5 ms. Observers maintained fixation throughout the trials.

The intensity of the spot was modulated linearly (verified by photometer measurements) by controlling the number of brief ($1.2 \mu\text{s}$) intensifications within each 5 ms interval. The absolute intensity of the stimulus was not measured, but was always calibrated to a given standard with a UDT model 40X photometer before each experimental session. Threshold intensity is reported in arbitrary logarithmic units as a sensitivity measure equal to $3 - \lg$ (number of intensifications at threshold).

All thresholds were measured under computer supervision with a multiple interleaved staircase (Cornsweet 1962). The observer responded to each trial (which was accompanied by an audible tone) by pressing one of two buttons, indicating whether or not he saw the stimulus. If he saw it, the intensity of that condition was decremented on the next trial, otherwise the intensity was incremented, first in 0.3 and then in 0.1 logarithmic steps. Thresholds, taken as the average intensity of the last 15 trials, were measured separately three times, in every case yielding a small standard error of the mean less than the size of the figure symbols. Six staircases, each of 32 trials, were run concurrently, one for each stimulus duration. Stimulus speed was constant within each trial block, but randomized between blocks within each experimental session.

The results of only one observer, myself (a well corrected hypermetrope), are reported here, but all major results have been verified on at least one other observer, all observers giving very similar results. As an additional precaution against possible subject bias, I have checked certain key data points with a two-alternative forced-choice procedure (see Burr (1980c) for details). In every instance the results of these measurements agreed almost perfectly with those of the less cumbersome yes-no staircase.

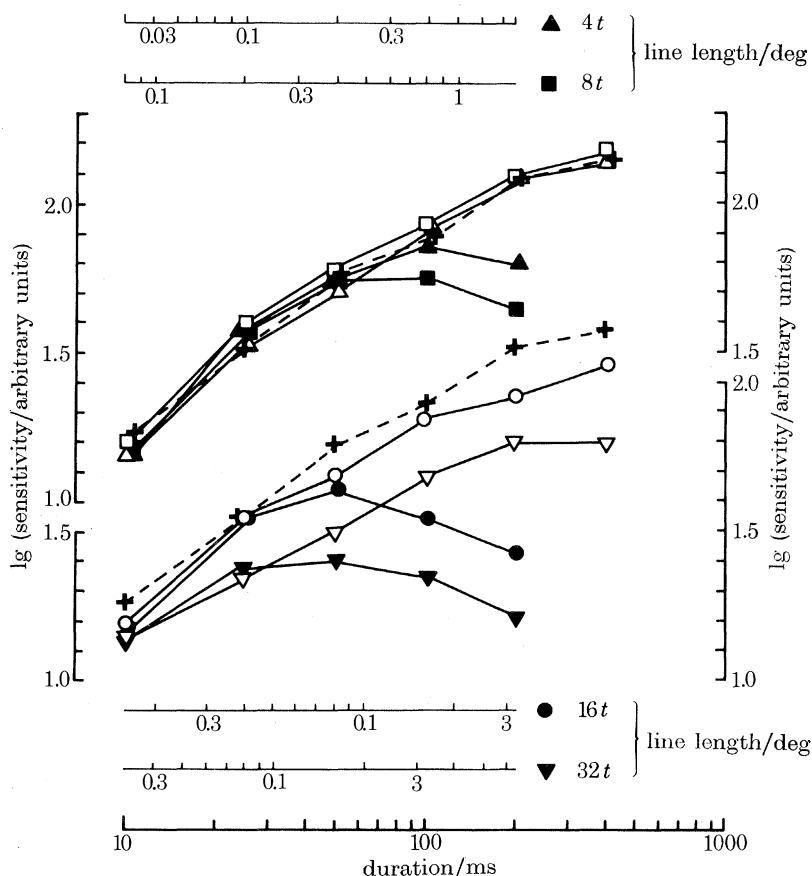


FIGURE 1. Sensitivity to moving point sources and stationary lines as a function of stimulus duration. To avoid clutter, the two curves are vertically divided into two groups, with the stationary point thresholds (+) plotted twice for comparison. The open symbols refer to points moving at 4 (Δ), 8 (\square), 16 (\circ) and 32 (∇) deg/s. The closed symbols of the same shape refer to stationary lines of length vt (where v is speed and t is duration), whose time- and space-averaged luminance corresponds to the moving stimulus. The separate abscissae indicate the line length for each condition. Note that these curves peel away from the others at 50 ms, long before summation is completed for the moving targets.

Moving dots

Results

Figure 1 summarizes the results of this and the next experiment. To avoid clutter the curves have been vertically separated into two groups, with the measurements of stationary point stimuli plotted twice for comparison. Consider

here only the data points depicted by open symbols, which represent the moving points.

The clear result of this experiment is that the thresholds for stimuli moving at speeds of up to 8 deg/s are virtually identical to those of stationary stimuli. In both complete summation occurs for up to about 100 ms, followed by partial summation for some time afterwards. Even at 16 deg/s summation is nearly complete, finally failing at 32 deg/s.

Thus the visual system is capable of summing energy from images in motion across the retina, provided that they do not move too quickly. This capacity is obviously of paramount survival importance, serving to help detect and recognize low-contrast targets in motion, our natural panorama being seldom completely stationary.

However, it is far from obvious how the summation is achieved. A target moving at 16 deg/s traverses 1.6° of retina during the period in which summation is occurring. Therefore, to be summated completely, it must be pooled from over this substantial region of retina. In other words, summation of a moving target implies summation over spaces as well as over time. However, the extent of spatial summation observed here would seem to be considerably greater than that previously reported. For example, Barlow's (1958) results show that total summation for a stationary light disk fails when its diameter exceeds about 0.5° .

Obviously, there are difficulties in comparing my results with those reported elsewhere, as the experimental conditions are far from identical; and it is well known that spatial summation varies greatly with factors such as luminance and stimulus duration. Therefore, I have remeasured spatial summation under conditions that mimic exactly those of the previous experiment.

Stationary lines

During its travels, a moving dot traverses a region of retina equal to the product of its image speed and exposure duration. Therefore, the appropriate comparison stimulus to a moving dot is a short stationary line of length equal to the product of the stimulus speed and exposure duration.

The methods were as just described except that the *X* axis of the oscilloscope was disconnected from the computer interface and driven with a fast (3 MHz) triangle wave from a Farnel FG2 waveform generator, to produce a line of the same energy (integrated over space) as the moving dots. The amplitude of the waveform was varied in direct proportion to the stimulus duration (by analogue multiplication under computer control), with the constant of proportionality chosen so that the length of the line corresponded to the distance traversed by a dot moving at a particular speed; that is, the length of these lines was such that the integral over time was the same as that for the moving dots. Thus the two stimuli have the same integrals over space and time, but differ in their spatio-temporal structure.

The results are depicted by the solid symbols of figure 1, superimposed on the previous results for ready comparison. Symbols of the same shape represent

stimuli of equal integrated length. The four separate abscissae show the line length (or distance travelled) of each condition.

Clearly, the combined temporal and spatial summation of the stationary lines is considerably less than that for a small target moving over a region of the same extent. For short durations the sensitivity curves peel away, and they actually fall for longer durations. Yet after this break energy from the moving targets continues to be summated almost totally for a further 50 ms.

As with the two flash experiments of Ikeda (1965), these results cannot be explained by temporal and spatial integration. Were integration the only process involved, then detectability of stimuli would be determined solely by the total energy of the stimulus (provided that the spatial and temporal extent does not exceed the integration limits). Thus distribution of energy within those regions would be irrelevant: integration destroys fine patterning information. Yet figure 1 shows that this is not so. Targets with identical spatial and temporal integrals yield widely different estimates of the extent of spatial and temporal summation. Clearly, some mechanism specialized for motion perception is involved.

A final observation about these experiments is that at image speeds of up to 16 deg/s the spot appeared single and sharp. It did not look like the smeared line corresponding to its spatial integral. This observation is pursued elsewhere (Burr 1980*a*).

SINUSOIDAL GRATINGS

To investigate further the summation of moving targets, and to attempt to account for the seemingly paradoxical results of the previous experiment, I have measured the temporal summation of the visual system with a different type of stimulus, drifting sinusoidal gratings.

There is now firm evidence that visual information is processed (initially at least) by a system of independent detectors or *channels*, each selectively tuned to a restricted range of spatial frequencies (see for example: Campbell & Robson 1968; Blakemore & Campbell 1969; Sachs *et al.* 1971; Graham *et al.* 1978). The channels differ markedly in their temporal response (Robson 1966): those tuned to high spatial frequency (say 20 cycle/deg) prefer low temporal frequencies, responding best to stationary gratings, and resolving only up to about 15 Hz, while those tuned to low spatial frequency (say 0.5 Hz) prefer higher temporal frequencies, responding best to signals of 5–10 Hz, attenuating both higher and lower frequencies, and resolving signals of up to 30–40 Hz.

The temporal summation also varies with spatial frequency. Barlow (1958) found that the critical duration of summation was longer for small (high spatial frequency) spots of light than for large (low spatial frequency) spots of light. More recently, Tolhurst (1975) has observed a similar effect with sinusoidal gratings: those of high spatial frequency summate for much longer than those of low spatial frequency.

Here I report measurements of contrast sensitivity for gratings of various spatial

frequency, both stationary and drifting at various temporal frequencies. They show that, provided that the gratings drift at the preferred temporal frequency, both high and low frequency gratings are summated over time by the visual system.

Methods

The general procedure was the same as that for the previous experiment: all thresholds were measured with a yes-no staircase, again key data points being checked with a forced-choice procedure.

Gratings were computed by the PDP-8/I computer and displayed on the face of a cathode ray oscilloscope at 150 frames per second, 1000 lines per frame, by means of the standard television technique of Schade (1956). The visible screen was a circle, 20 cm in diameter, of mean luminance 200 cd/m², surrounded by a 1 m² mask of the same luminance. As the viewing distance varied between experiments, screen size in degrees of visual angle is noted below each figure.

To confine temporal frequency to a narrow band, gratings should ideally be of infinite, or at least long, duration. As this is clearly not possible in the present experiment (which measures the effect of duration) the gratings were multiplied by a *raised cosine* temporal envelope, $m = \frac{1}{2}[1 - \cos(2\pi t/\tau)]$, where m is contrast, t time and τ total stimulus duration. As this closely approximates a Gaussian waveform it ensures minimal spread of frequencies at any given duration (see Bracewell 1965). Exposure duration is taken as the width of the temporal envelope at half height (i.e. $\frac{1}{2}\tau$).

Contrast was varied between trials with a computer-controlled attenuator, first in steps of 0.3, then 0.1, logarithmic unit. As usual, contrast sensitivity is defined as the inverse of the peak contrast required for threshold, i.e. $(L_{\max} + L_{\min}) / (L_{\max} - L_{\min})$.

Results

Figure 2 shows the results of threshold measurements for gratings of relatively large extent (at least six cycles) of high, medium and low spatial frequency, drifting at a range of temporal frequencies. The results are summarized by three separate families of curves, one for each spatial frequency.

Consider first the results for stationary gratings. These essentially replicate those of Tolhurst (1975), showing summation for up to nearly 200 ms for high frequency gratings, but to only 50 ms for medium and low spatial frequencies. Indeed, the 0.5 cycle/deg grating actually shows negative summation after 50 ms. Sensitivity steadily decreased with longer durations, to yield thresholds twice as high at 400 ms as at 50 ms exposure.

The trend for drifting gratings, however, is quite different. At low spatial frequencies, image motion has the effect of greatly increasing the duration of summation to such an extent that at the optimal drift rates of 5 and 10 Hz the summation time is about as long as for high frequency stationary gratings. At high spatial frequencies the reverse holds: summation time is shorter at the faster drift rates.

Thus the summation time depends on both the temporal and the spatial properties of the stimulus. High frequency gratings summate well when stationary but poorly when drifting, whereas low frequencies summate rather poorly, in fact negatively, when stationary, but well when drifting, indeed as well as the high frequency stationary gratings.

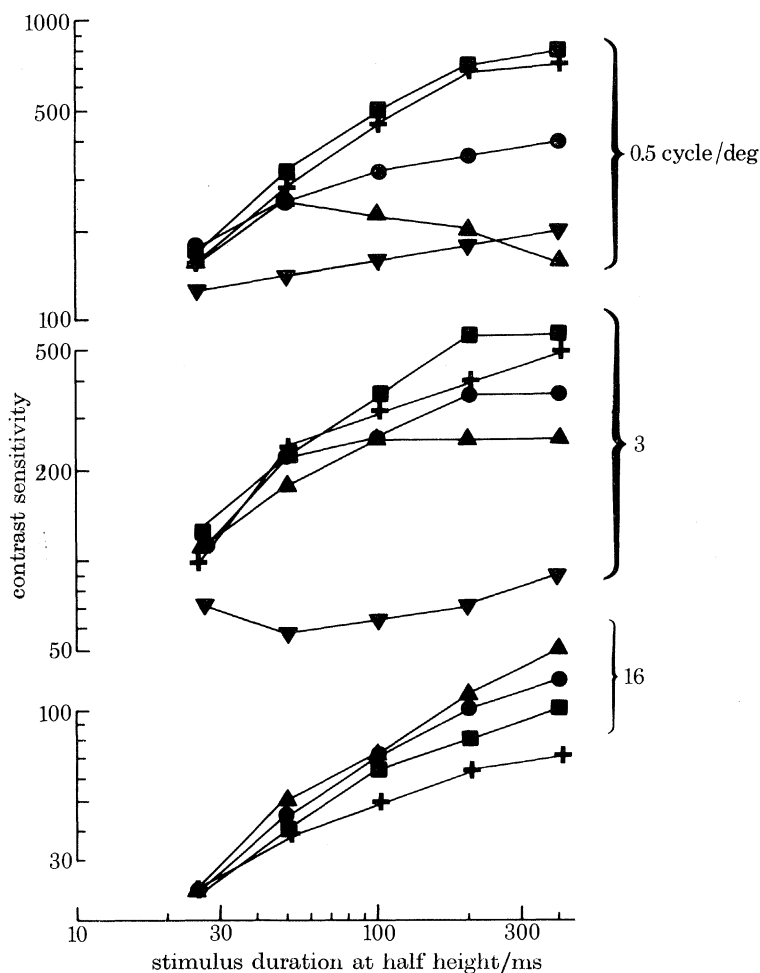


FIGURE 2. Contrast sensitivity to drifting and stationary gratings as a function of duration. The three families of curves (0 (▲), 1.5 (●), 5 (■), 10 (+), 20 (▼) Hz) are measurements of gratings of spatial frequency 0.5, 3 and 16 cycle/deg. The screen size varied with spatial frequency, being 12, 2 and 0.7 deg respectively. Note that the 16 cycle/deg gratings summate well when stationary, but steadily less well when drifting. However, at 0.5 cycle/deg the reverse is true, the 5 and 10 Hz gratings summing best and the stationary ones failing to summate at all after 50 ms. Indeed, these low spatial frequency gratings actually become less visible at durations longer than 50 ms.

ANALOGUE FILTER SIMULATIONS

Obviously, these results cannot be explained by the classical integration model. The temporal integral of a grating drifting at 10 Hz is zero at 50, 100, 150 and 200 ms (duration taken at half height), yet a 0.5 cycle/deg grating continues to summate up to 200 ms. Similarly, integration could never account for the negative summation of the stationary coarse grating.

As mentioned earlier, an alternative approach is to describe the visual system as a linear filter, performing both integration and differentiation, followed by some threshold device. If a system is linear, its response to signals of any frequency and duration can be predicted from the frequency transfer function. In practice, the transfer function cannot be determined for the visual system without making the (probably unreasonable) assumption of constant gain. But for the purposes of this study, it is sufficient to know the transfer function at threshold, which is specified by the temporal contrast-sensitivity function (with the additional assumption of linear phase), the minimum amplitude of temporal modulation required to see flicker for each flicker frequency (Robson 1966).

Previous researchers (e.g.: Sperling & Sondhi 1968; Kelly 1971; Roufs 1972; Watson & Nachmias 1977) have related the contrast-sensitivity function to the summation results computatively by means of Fourier transform of the transfer function, which yields the *impulse response function*, which can in turn predict the response to multiple or continuous signals. However, another approach, which is perhaps less mystifying to those not fully acquainted with the intricacies of Fourier theory, is direct analogue simulation. To examine whether the results of figure 2 can be predicted from the temporal contrast-sensitivity measurements, I constructed electronic filters to the specification of the human visual system (as determined by Robson (1966)†), and measured the magnitude of their response to signals of the same frequency and duration as those used for the experiment of figure 2. The simulations show that, with the additional assumption of probability summation, the summation results are well predicted by the contrast-sensitivity measurements.

Methods

As noted earlier, the temporal contrast-sensitivity function varies considerably with spatial frequency, resembling a bandpass filter at low spatial frequencies and a low pass filter at high spatial frequencies. I therefore constructed two separate filters, one to simulate the 0.5 cycle/deg 'channel', the other to simulate the 16 cycle/deg 'channel'. The 0.5 cycle/deg channel consisted of two electronic

† Robson's measurements were made with counterphase-modulated rather than drifting gratings. However, a counterphase grating is physically equivalent to two drifting gratings of half amplitude moving in opposite directions; and indeed Levinson & Sekuler (1975) have shown that the contrast-sensitivity curves for the two types of modulation are of the same shape, with sensitivity to drifting gratings being predictably twice as good at all frequencies. Thus, for the purpose of this simulation, Robson's measurements are sufficient.

filters connected in series to produce a bandpass filter, a high pass filter attenuating the low frequencies at 6 dB per octave, and a low-pass filter attenuating the high frequencies more steeply at 24 dB per octave. For the 16 cycle/deg channel, only a 16 dB per octave low-pass filter was used.

The transfer characteristics of those filters were examined by passing a sine wave of exponentially increasing frequency (from a Farnell FG 1 function generator) through the filters, taking the logarithm (by a Hewlett-Packard 7561A logarithmic convertor) and displaying the result on a storage oscilloscope whose gain was adjusted to display equal logarithmic axes. The filters were then trimmed to produce a response curve of approximately the same shape as that measured psychophysically by Robson. These curves are shown in figure 3.

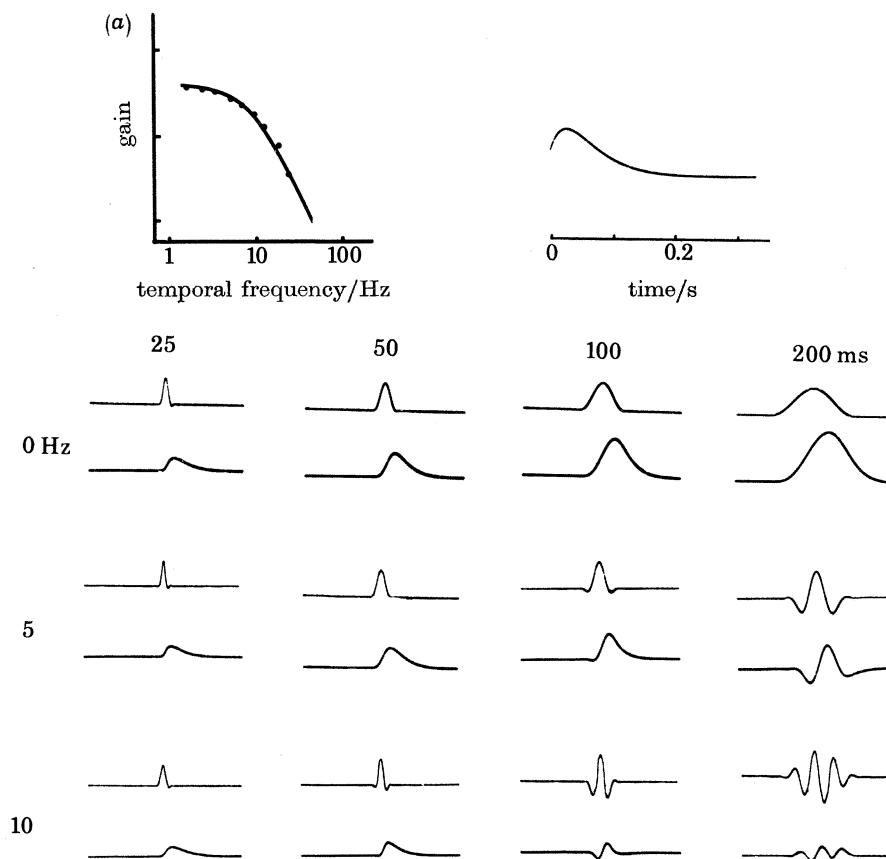


FIGURE 3. For description see opposite.

The waveforms of the simulation were sinusoidal, multiplied by a raised cosine envelope. This would be the luminance distribution over time at a particular retinal location. One problem with this approach is to choose the starting phase of the waveform, as in the actual experiment (with drifting gratings) it varies with retinal position. The solution was to assume that detection takes place at the site

Temporal summation of moving images

331

producing maximum response. The measurements of the strength of the filter output reported in figure 4 are those made with the phase of the input waveform adjusted to produce maximum output.

Results

Photographic records of the filter output are shown in figure 3. The upper traces of each figure show the frequency response and the impulse response functions of each filter (the latter produced by recording the response to a brief rectangular pulse).

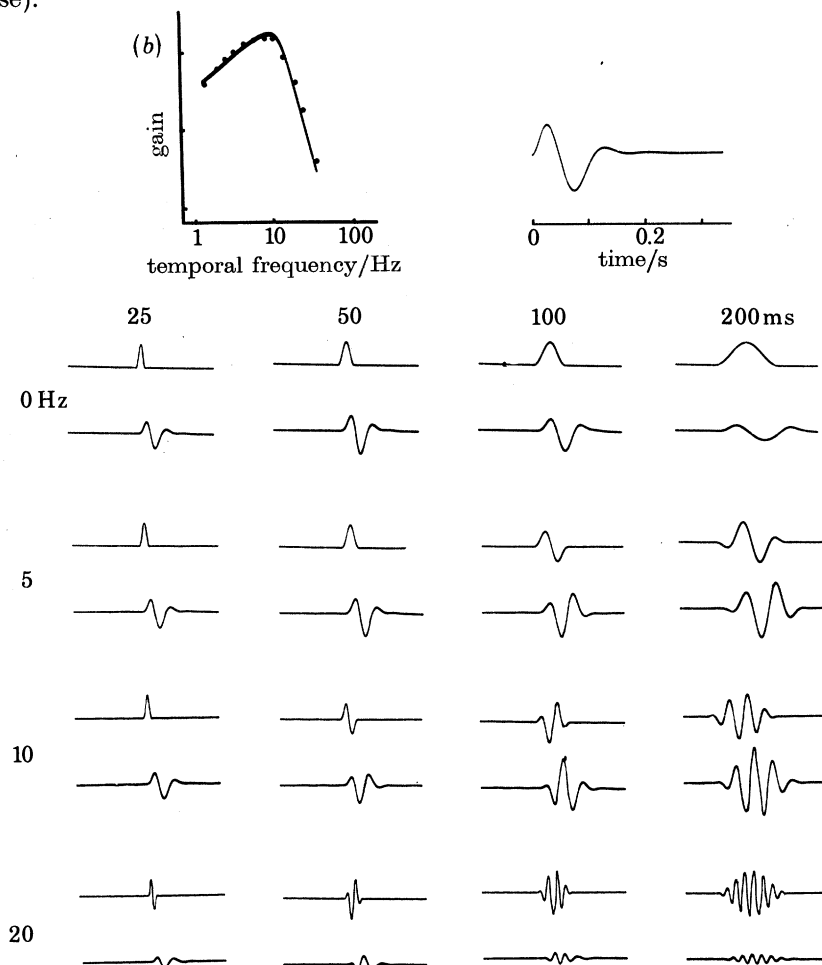


FIGURE 3. Photographic records of the output of the simulated visual channels. The upper trace of each pair is the input signal, the lower the filter response. The top traces show the frequency response function of each channel superimposed on Robson's (1966) measurements of contrast sensitivity, together with the impulse response functions. The impulse response function of the 16 cycle/deg channel (*a*) is monophasic, implying integration and hence strong summation of only stationary or near-stationary stimuli. On the other hand, the impulse response function of the 0.5 cycle/deg channel (*b*) is diphasic, reflecting resonance at about 8 Hz. It therefore strongly summates only those stimuli that are modulated at or near that frequency.

Note first the difference in impulse response functions. Whereas the 16 cycle/deg channel is monophasic, exhibiting a *sustained*, steadily decaying response, the 0.5 cycle/deg channel is diphasic, a *transient* response immediately followed by a negative response of the same strength. This basic difference in response to an impulse leads to differences in response to stimuli of varying duration and frequency.

The remaining traces show the response to modulated stimuli. The upper trace of each pair is the input signal, the lower trace the filter response. Consider first the 16 cycle/deg channel. At brief durations (25 ms at half height) the response to all frequencies is essentially the same, as indeed they are virtually the same stimulus. The effects of the filter become more pronounced at longer durations. It resonates to low temporal frequencies, becoming longer and larger as the stimulus is extended over time. However, it attenuates the high frequencies. The attenuation occurs because waveforms of limited extent or duration contain not only the nominal frequency, but a spread of frequencies whose width varies inversely with the duration. Any brief waveform, therefore, has a broad temporal spectrum including both high and low frequencies. Thus a brief waveform of nominally high temporal frequency has low frequencies which are passed by the filter. At 200 ms the stimulus is restricted to a narrow band around 10 Hz, which the filter attenuates strongly.

Similarly the 0.5 cycle/deg channel shows decreased response to longer-duration high temporal frequencies, but not until 20 Hz, as the filter has a higher high frequency cut. The response to 5 and 10 Hz waveforms actually increases up to 200 ms. But again, this is easily explained in frequency space. The bandpass filter is tuned to pass only a narrow range of frequencies around 6 Hz or so. As we have seen, brief exposures contain a broad range of frequencies, much broader than the width of the filter, so that only a fraction is passed. The filter responds best to stimuli of several periods, which contain a fairly narrow band of frequencies falling within its operating range.

Note also the response to the 0 Hz waveform. Like the psychophysical measurements, the response decreases with duration after 50 ms. This results from differentiation characteristics of high pass filters; or, if one considers the frequency content, the long pulse contains only very low frequencies, to which the filter does not respond.

Figure 4 summarizes the results of the simulation, with the filter gain (in arbitrary logarithmic units) plotted against duration. The psychophysical measurements of figure 2 are replotted for comparison.

The results are in good qualitative agreement; both the simulations (closed symbols) and the psychophysical measurements (open symbols) yield similar sorts of curves. The general shape is the same, but the absolute slopes are different. Notice that the difference is not random: in most cases the psychophysical curves are steeper (more positive slope) by about 25%, the amount expected from probability summation. Stimuli of long duration produce not only a stronger but also a longer response, which therefore has a greater probability of being detected by the noise-perturbed detection process. Exhaustive studies (see, for example:

Temporal summation of moving images

333

Sachs *et al.* 1971; Watson, 1979) show that the predicted improvement is a doubling in sensitivity for every four doublings in duration, an increase in slope of 25 %.

Thus, after probability summation is taken into account, the curves match each other quite well. That is to say, the summation times of drifting sinusoidal-modulated gratings of any given spatial frequency may be accounted for by a single linear temporal filter.

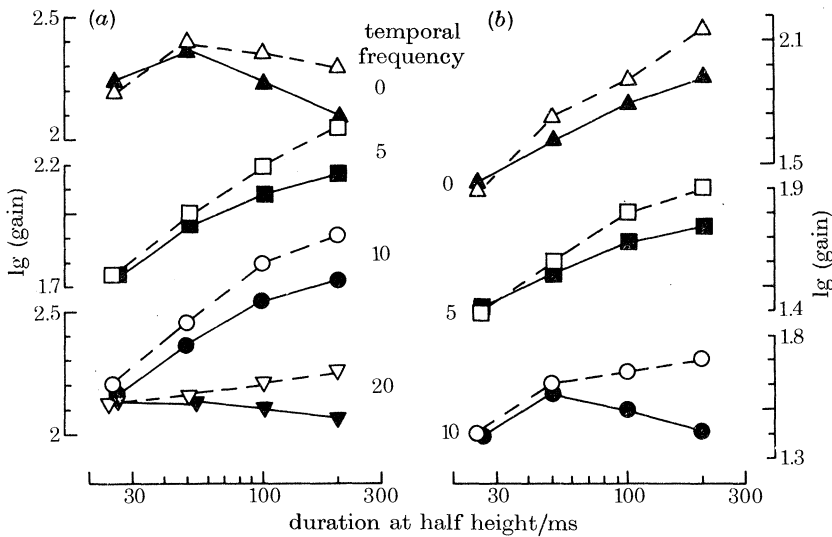


FIGURE 4. Summary of the analogue filter simulation results (closed symbols), compared with the psychophysical results of experiment 2 (open symbols); (a) 0.5 cycle/deg; (b) 16 cycle/deg. To avoid clutter, the curves are separated into six families, each with its own ordinate. The measurements at 5 Hz, 25 ms, are restrained so that the simulations coincide with the contrast-sensitivity data, but all other points were free to vary. After probability summation is taken into account, the simulations predict the psychophysical results quite closely, implying a single linear temporal channel for each spatial frequency.

SMALL GRATING PATCHES

Moving targets, both dots and sinusoidal gratings, are summated over time by the visual system (figures 1, 2). Neither result can be explained by combined temporal and spatial integration. The grating results, however, are all well accounted for by assuming a single linear temporal filter (for each size-specific spatial channel), which is followed by some perturbed threshold device (figures 3, 4). The question that now arises is whether this filter model can contribute anything towards the explanation of the results with single light points. That is, do results obtained with sinusoidal gratings generalize to more natural conditions, where a single, spatially localized target is in motion?

A sinusoidal grating drifts continuously over one retinal region during the entire

presentation period, thereby stimulating continuously all retinal sites within that region. One can therefore consider the waveform at any one arbitrary site, where it will be modulating sinusoidally with time. On the other hand, a point of light (or indeed any naturally moving target) has a localized image of finite size, which moves successively from one retinal site to another, stimulating each only momentarily. Therefore, to investigate the response over time to such a stimulus, it is necessary to consider the response of not just one retinal site, but a whole region of retina, the region over which the image traverses during one period of temporal summation. In other words, complete summation of signals from a moving point requires not only summation over time, but also summation over space.

This experiment, therefore, measures the interaction between spatial and temporal summation during detection of a small localized target, to see whether the filter model can be extended to this type of target. The stimulus was a small 'patch' of grating (a sinusoidal grating smoothly vignettted by a raised cosine envelope), caused to drift across the visual field. This is a stimulus that combines the virtues of both a point and a grating: it has a finite size and spatial location, and therefore traverses many retinal sites, while at the same time allowing for independent control of both spatial and temporal frequency. Moreover, because the space averaged luminance is constant over time and the time averaged luminance is constant over space, the pattern leaves no residual stationary pattern.

A problem with the measurements with dots is that the temporal frequency of the stimulus is markedly different when the dot is stationary from when it is moving. Here it is possible to compare the detectability of a moving stimulus with that of a stationary one, without varying the temporal frequency. In both conditions the grating drifted at 10 Hz, but in one the grating drifted within its vignetting window (which itself remained stationary), while in the other the window drifted along with the grating at the same velocity. Thus for one case the same small retinal site was continuously stimulated and for the other the stimulus moved over many sites.

Methods

The methods were identical to those of the previous experiment except that the waveform was vignettted to only one cycle, by multiplication with a single cycle of raised cosine waveform of the same frequency. In one condition this envelope remained stationary while the grating drifted within it; in the other it drifted at the same velocity as the grating, so that the relative phase between it and the grating was always zero (see inset in figure 5). The two conditions were randomly interleaved within each session.

Measurements were made at four spatial frequencies†, ranging from 0.05 to

† It should be borne in mind that, as there was only one cycle of grating present, there was considerable spread of spatial frequencies around the nominal frequency. However, this does not greatly affect the experiment.

Temporal summation of moving images

335

20 cycle/deg, at velocities ranging from 0.5 to 200 deg/s. The temporal frequency of drift was always 10 Hz, chosen as the maximum frequency for strong temporal summation. The height of the bars was always equal to four cycles of grating.

Results

The results of the measurements are summarized in figure 5. Clearly, there is little difference between the two experimental conditions: both the moving patch (\blacktriangle) and the stationary patch (\bullet) show strong summation at least up to 100 ms. Here the moving patch is only slightly less visible than the stationary one \dagger , although it must traverse a region of retina equal to about 2 cycle of grating. Thus the translation *per se* of an image across retina does not impair its visibility.

Detection of the moving grating patch requires not only 100 ms of temporal

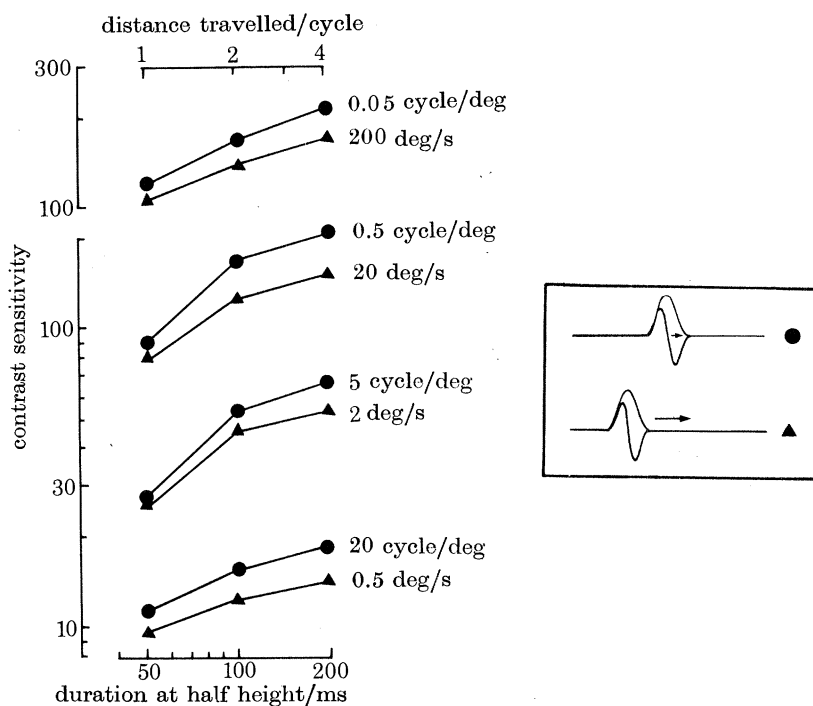


FIGURE 5. Comparison of contrast sensitivity for one cycle of sinusoidal grating drifting *within* its vignetting window (\bullet) with one drifting *along with* the window at the same velocity (\blacktriangle), both measured as a function of exposure duration. The upper abscissa shows the distance traversed by the moving window, expressed as the number of cycles of grating. Note that detectability and summation is about the same for the two conditions, being only about 2 dB higher for the stationary window than for the moving one.

\dagger It is possible that this small difference in visibility can be accounted for by the difference in phase of the two conditions: the drifting patch was always in zero phase relative to the grating, whereas the phase relationship for the stationary patch varied continuously, causing a variation in amplitude of the compound, and hence a greater peak amplitude than when the two waveforms are combined in zero phase.

summation, but also spatial summation of about 2 cycle of grating. That is to say, the catchment area, or receptive field of the detector responding to the target must extend over a region of about 2 cycle, regardless of the spatial frequency. This is sufficiently large to summate over space most of the energy that is summated over time, for a target drifting at the maximum frequency for efficient temporal summation (10 Hz). The regions of temporal and spatial summation seem to be matched to optimize the combined summation over time and space, which is essential for the efficient detection of a naturally moving target.

The relevance of these results is that they show that there is little difference between the temporal summation of a target drifting within a window, and hence not translating in position, and one that translates smoothly over time, following a motion trajectory. For stimuli of comparable size and duration, visibility thresholds are completely determined by the temporal and spatial frequency of the stimulus. Thus results obtained with grating of large extent, and those of the analogue simulations (which modelled only temporal frequency, neglecting space), are probably also applicable to small targets that traverse a region of retina.

How then can the filter model account for the results of figure 1, that small moving dot targets are summated over time (and space) to a much greater extent than are comparable line stimuli? As stated earlier, there is now mounting evidence that initial detection of threshold stimuli is achieved by a system of independent channels selective for spatial frequency (see Braddick *et al.* (1978) for a review). These spatially selective channels have different temporal frequency characteristics (Robson 1966), which has been shown in this investigation to lead to qualitative differences in temporal summation of different spatial frequencies.

In Fourier space, a dot is the sum of all travelling cosine waves, equally weighted on a linear scale. That is to say, every spatial and temporal frequency is represented in equal proportions. Thus a dot can be detected by any spatial-frequency channel and at threshold will presumably be detected by the most sensitive. Robson's (1966) results show that, for stationary targets, the high-medium range channels are the most sensitive; as these channels summate over time for about 100 ms, they can account for the temporal summation of stationary dots. When the dot is caused to drift, the 'component cosinusoids' become modulated over time. Now the lower frequency channels become the more sensitive (Robson 1966; Burr 1980*b*) and presumably detect the dot; these channels also summate over time for about 100 ms, accounting for the temporal summation of moving dots. The only stimulus that will not be summated well by this system of channels is a stationary line: a line is too large to be spatially summated by the channels selective for high spatial frequencies (see also: Howell & Hess 1978; Robson & Graham 1980), and, because the energy is not modulated over time, it will not be temporally summated by the selective for low spatial frequencies channels. Thus the results of figure 1, which are quite puzzling when considered solely in terms of spatial and temporal integration, are readily accounted for by a simple model that considers each spatially selective channel to be a linear temporal filter.

DISCUSSION

The natural visual environment is dynamic. One seldom views a totally stationary scene, and, typically, the objects in motion are those of greater interest, often of great importance to survival. If these moving stimuli are to be detected and resolved with the same efficiency as stationary ones, there must exist in vision a system capable of summing over time the signals emanating from their images. Without such summation sensitivity would inevitably suffer.

The evidence of this investigation favours the existence of such a system. There seems to be a set of visual channels tuned to a range of relatively low spatial frequencies and high temporal frequencies that are specialized to detect and analyse targets in motion. By virtue of their temporal tuning, these channels summate over time the signals of moving targets, thereby fortifying them for transmission through the inherently noisy visual system.

Supplementing this system, there exists another set of channels tuned to higher spatial frequencies and lower temporal frequencies. This system does not respond well to moving stimuli. Indeed, there is evidence to suggest that the channels that are tuned to the highest spatial frequencies are completely incapable of signalling information about motion (Campbell & Maffei 1979).

Of course, the idea of functionally separate channels for motion is not entirely new. Eight years ago Tolhurst first proposed that motion and pattern information are analysed separately by independent channels (Tolhurst 1973; Kulikowski & Tolhurst 1973). The temporal and spatial properties of these channels agree well with those observed in the present study. However, despite the strong general agreement, a fundamental disagreement should be clarified.

This disagreement lies chiefly in the interpretation of the roles assigned to the two systems. Tolhurst asserts that 'the movement-dependent channels alone carry information leading to the sensation of movement. The movement independent channels might be largely responsible for the analysis of the spatial structure of the stimulus, whether moving or stationary' (Tolhurst 1973). That is, the motion channels signal only information about the motion *per se*, or velocity of the stimulus, contributing nothing to its spatial analysis.

This study supports strongly Tolhurst's idea of two functionally separable sets of visual channels, distinguishable on the basis of their temporal and spatial tuning. However, it does not support the idea that these channels serve to analyse separately two aspects of the one stimulus, one set signalling only velocity information and the reporting on spatial form. Rather, it has been shown that the temporal tuning of the motion channels not only serves to enhance motion sensitivity, but also renders them an ideal mechanism for the analysis of the spatial structure of the target in motion. The temporal tuning leads to temporal summation, which in turn substantially improves signal to noise levels. The signal emerging from these channels will be far stronger and more reliable than that of the movement-independent channels. It seems implausible that the spatial information should be

extracted by the movement-independent channels, which, lacking the capacity for temporal summation of moving stimuli, can provide only a very weak signal. Both their spatial and temporal properties render them relatively impotent to transmit any sort of information arising from moving stimuli, be it about motion or about spatial structure. Rather, these high spatial frequency 'sustained', channels probably serve to complement the movement system by performing a detailed spatial analysis of stationary, or slowly moving, objects.

I am highly indebted to Denis Pelli for considerable support at all stages of this research. The valuable help of Fergus Cambell, Horace Barlow and many other colleagues of the Craik Laboratory is also gratefully acknowledged. Finally I thank the Hackett Foundation of the University of Western Australia for funding me during the experimentation stages, and the Royal Society of London for a European Science Exchange Programme Fellowship during the preparation of this paper.

REFERENCES

- Barlow, H. B. 1958 Temporal and spatial summation in human vision at different background intensities. *J. Physiol., Lond.* **141**, 337–350.
- Blakemore, C. & Campbell, F. W. 1969 On the existence of neurones in the visual system selectively sensitive to the orientation and size of retinal images. *J. Physiol., Lond.* **225**, 437–445.
- Bloch, A. M. 1885 Experience sur la vision. *C. r. Séanc. Soc. Biol.* **37**, 493–495.
- Bracewell, R. 1965 *The Fourier transform and its applications*. New York: MacGraw-Hill.
- Braddick, O. J., Campbell, F. W. & Atkinson, J. 1978 Channels in vision: basic aspects. In *Perception. Handbook of sensory physiology*, vol. 3. Berlin: Springer-Verlag.
- Burr, D. C. 1980a Motion smear. *Nature, Lond.* **284**, 164–165.
- Burr, D. C. 1980b Sensitivity to spatial phase. *Vision Res.* **20**, 391–396.
- Burr, D. C. 1980c Contrast sensitivity at high velocities. (In preparation.)
- Campbell, F. W. & Maffei, L. 1979 Stopped visual motion. *Nature, Lond.* **278**, 55.
- Campbell, F. W. & Robson, J. G. 1968 Applications of Fourier analysis to the visibility of gratings. *J. Physiol., Lond.* **197**, 551–566.
- Cornsweet, T. N. 1962 The staircase method in psychophysics. *Am. J. Psychol.* **73**, 483–491.
- De Lange, H. 1954 Relationship between critical flicker-frequency and a set of low frequency characteristics of the eye. *J. opt. Soc. Am.* **44**, 380–389.
- Graham, C. H. & Margaria, R. 1935 Area and intensity-time relation in the peripheral retina. *Am. J. Physiol.* **113**, 299–305.
- Graham, N., Robson, J. G. & Nachmias, J. 1978 Grating summation in fovea and periphery. *Vision Res.* **18**, 815–826.
- Howell, E. R. & Hess, R. F. 1978 The functional area for summation to threshold for sinusoidal gratings. *Vision Res.* **18**, 369–374.
- Ikeda, M. 1963 Study of interrelations between mechanisms at threshold. *J. opt. Soc. Am.* **53**, 1305–1313.
- Ikeda, M. 1965 Temporal summation of positive and negative flashes in the visual system. *J. opt. Soc. Am.* **55**, 1527–1534.
- Kelly, D. H. 1971 Theory of flicker and transient response. *J. opt. Soc. Am.* **61**, 537–546 and 632–640.
- Kulikowski, J. J. & Tolhurst, D. J. 1973 Psychophysical evidence for sustained and transient neurones in the human visual system. *J. Physiol., Lond.* **232**, 149–162.

Temporal summation of moving images

339

- Levinson, E. & Sekuler, R. 1975 The independence of channels in human vision selective for direction of movement. *J. Physiol., Lond.* **250**, 347–366.
- Rashbass, C. 1970 The visibility of transient changes in luminance. *J. Physiol., Lond.* **210**, 165–186.
- Robson, J. G. 1966 Spatial and temporal contrast sensitivity function of the visual system. *J. opt. Soc. Am.* **56**, 1141–1142.
- Robson, J. G. & Graham, N. 1980 Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Res.* (In the press.)
- Roufs, J. A. J. 1972 Dynamic properties of vision. II. Theoretical relationship between flicker and flash thresholds. *Vision Res.* **12**, 279–292.
- Sachs, H. B., Nachmias, J. & Robson, J. G. 1971 Spatial frequency channels in human vision. *J. opt. Soc. Am.* **61**, 1176.
- Schade, O. H. 1956 Optical and photoelectric analogue of the eye. *J. opt. Soc. Am.* **46**, 721–739.
- Sperling, G. & Sondhi, M. M. 1968 Model for visual luminance discrimination and flicker detection. *J. opt. Soc. Am.* **58**, 1133–1145.
- Tolhurst, D. J. 1973 Separate channels for the analysis of the shape and the movements of a moving visual stimulus. *J. Physiol., Lond.* **231**, 385–402.
- Tolhurst, J. D. 1975 Reaction times in the detection of gratings by human observers: a probabilistic mechanism. *Vision Res.* **15**, 1143–1149.
- Watson, A. B. 1979 Probability summation over time. *Vision Res.* **19**, 515–522.
- Watson, A. B. & Nachmias, J. 1977 Patterns of temporal interaction in the detection of gratings. *Vision Res.* **17**, 893–902.