

Critical band masking in optic flow

PETER J. BEX¹, STEVEN C. DAKIN², & ISABELLE MARESCHAL²

¹Division of Visual Rehabilitation Research, The Institute of Ophthalmology, London, UK, and ²Division of Visual Sciences, The Institute of Ophthalmology, London, UK

(Received 30 November 2004; revised 1 July 2005; accepted 5 July 2005)

Abstract

Visual processing has been widely investigated with narrow band stimuli at low contrasts. We used a masking paradigm to examine how visual sensitivity under these conditions compares with the perception of the direction of heading in real scenes (i.e., with dynamic natural images at high contrasts). We first confirmed and extended previous studies showing biases in the amplitude distribution for spatial frequency, temporal frequency, speed and direction in dynamic natural movies. We then measured contrast thresholds for identification of the direction of motion for an observer traveling at various speeds. In spite of differences in contrast sensitivity and large non-uniformities in the amplitude content of the stimuli, contrast thresholds were relatively invariant of spatial frequency and completely invariant of temporal frequency, speed and direction. Our results suggest that visual processing normalises responses to supra-threshold structure at different spatial and temporal frequencies within natural stimuli and so equates their effective visibility.

Keywords: *Optic flow, masking, motion, natural images, driving*

Introduction

Human visual perception uses spatio-temporal variations of light falling on the retinae derive functional descriptions of the environment. Much of our understanding of this process, encapsulated in widely accepted *channel theories* of early vision (Blakemore & Campbell 1969; Campbell & Robson 1968; Graham & Nachmias 1971) comes from experiments with isolated, static sinusoidal gratings presented briefly at threshold contrast levels. Unlike experimental stimuli, natural images contain broad distributions of (among other things) spatial frequency, temporal frequency, orientation, speed and direction. Although random dot noise images are also commonly used in psychophysical studies and have a broad spectrum, the amplitude spectrum of white noise is quite unlike that of most natural images. While the amplitude of white noise is on average constant across spatial frequency, orientation and temporal frequency (for dynamic stimuli), the amplitude of natural scenes falls with spatial frequency with a characteristic slope of around -1.5 on log-log axes (Kretzmer 1952):

$$\text{amplitude}(f) = cf^{-\alpha} \quad (1)$$

where amplitude is averaged across all orientations, c is a constant, f is spatial frequency and α represents the slope on log-log axes. The value of α lies within a fairly narrow range

Correspondence: P. J. Bex, Division of Visual Rehabilitation Research, The Institute of Ophthalmology, 11-43 Bath Street, London EC1V 9EL, UK. Tel: 0207 608 4015. Fax: 0207 608 6983. E-mail: p.bex@ucl.ac.uk

(0.7–2.5) in achromatic images (Field 1987; Burton & Moorhead 1987; Hancock et al. 1992; Tolhurst et al. 1992; Ruderman 1994; Billock 1996; van der Schaaf & van Hateren 1996; van Hateren & van der Schaaf 1998). A comparable reduction in the amplitude of higher temporal frequencies has been reported for dynamic natural scenes and movies (Dong & Atick 1995a; van Hateren 1997; Billock et al. 2001). With respect to orientation, several studies have reported that horizontal and vertical structure is relatively higher in amplitude than oblique structure (Switkes et al. 1978; Hancock et al. 1992; van der Schaaf & van Hateren 1996; Coppola et al. 1998; Keil & Cristobal 2000; Hansen et al. 2003; Betsch et al. 2004).

Human sensitivity to static sinusoidal modulations in luminance has a classic inverted “U” shape, peaking at around 2–4 c/deg (Campbell & Green 1965; Campbell & Robson 1968). Sensitivity to temporal modulation of sine gratings depends on spatial frequency, in that lower spatial frequencies are more visible at higher temporal frequencies (Robson 1966; Kelly 1971; Kulikowski & Tolhurst 1973; Tolhurst et al. 1973). Contrast sensitivity is maximum for a drifting 3 c/deg grating stimulus at approximately 5 Hz, the so-called optimal motion stimulus (Watson & Turano 1995). Speed can be expressed as:

$$\text{Speed} = \text{TF}/\text{SF} \quad (2)$$

where TF is temporal frequency and SF is spatial frequency. This relationship produces a roughly flat peak contrast sensitivity for speed/velocity for the optimal combination of spatial and temporal frequency (Burr & Ross 1982).

Estimates show that channel bandwidths are approximately log scaled (Campbell et al. 1969; Maffei & Fiorentini 1973; Ikeda & Wright 1975; Movshon et al. 1978; DeValois et al. 1982), although bandwidths narrow to some extent at higher spatial frequencies (Blakemore & Campbell 1969; Tolhurst & Thompson 1981; DeValois et al. 1982). A straight line on a log-log plot of image amplitude versus spatial frequency indicates constant energy within bands whose width remains constant expressed in octaves. This means that the response of the visual system to natural images should be relatively constant across channels (Brady & Field 2000; Field 1987). Therefore, a low contrast natural image should first become detectable when its 3–4 c/deg components reach their detection threshold. If the pattern is dynamic, detection will first be possible when the 5 Hz component at 3–4 c/deg reaches its contrast detection threshold. To a first approximation (with static images) this speculation has been confirmed with a number of compound images (Watson 2000) and with natural scenes (Bex & Makous 2002).

At supra-threshold levels, the dependence of appearance on contrast sensitivity breaks down and perceived image contrast is relatively independent of spatial frequency (Bryngdahl 1966; Watanabe et al. 1968; Blakemore et al. 1973; Georgeson & Sullivan 1975; Kulikowski 1976; Cannon 1979; Bowker 1983; St John et al. 1987; Brady & Field 1995), a phenomenon termed *contrast constancy*. However, the supra-threshold apparent contrast of notch filtered natural scenes does dip when components at around 4 c/deg are removed (Bex & Makous 2002). This discrepancy—being more characteristic of performance near contrast detection-threshold—compromises our ability to predict the appearance of broad-band natural images from the visibility of low-contrast grating patterns.

Contrast constancy therefore suggests that, at supra-threshold contrasts, the responses of different visual channels are to some extent normalised and therefore that image components over 4 octaves (between 1–16 c/deg) are of approximately equal visibility. Several studies have examined how the visual system utilises these equally visible components to perform discriminations among high contrast complex images of letters or faces Peli et al. 1994; Solomon & Pelli 1994; Gold et al. 1999). For example, Solomon and Pelli (1994) measured threshold contrast for letter identification in the presence of low-pass and high-pass filtered

noise. Contrast sensitivity was highest when masks did not contain components at around 4–8 cycles per letter and was lowest when these components were present in the noise masks. These results were relatively invariant of the size of the letters, suggesting that contrast sensitivity *per se* does not limit performance in visual identification tasks at high contrasts. Analogous results have been reported for face identification, with optimal components for this task falling between 8 and 16 cycles per face (Gold et al. 1999; Peli et al. 1994).

If at supra-threshold contrasts, the responses of the visual system are relatively invariant of spatial scale, then the contrast sensitivity function can no longer predict visual performance. It might therefore be expected that the well-documented biases in the spectra of natural images come to determine visual sensitivity. We test this conjecture in the present study. We examine how the visual system integrates information across wide ranges of spatial frequency, temporal frequency, speed and direction in dynamic natural images. We adapt a noise masking paradigm to measure sensitivity to the direction of heading with dynamic natural images generated by self motion, known as optic flow (Gibson 1966; Koenderink 1986). We choose this task both because of its ecological relevance, and because it could be supported by image structure at a variety of spatial and temporal frequencies, allowing us to investigate interactions among them.

Methods: Stimuli

Natural scenes

Movies were recorded with a Panasonic NV-GS200EB digital video camera with no digital or optical zoom. Movies were recorded at 3 different speeds of self-motion: driving at 70 mph along a 2×3 lane highway; driving at 30 mph along a 2×1 lane road; walking along an urban street (see Figure 1 for examples of single frames from our movies). In practice, we found that a hand-held camera was too unstable to record the optic flow of a pedestrian and that steadier images were obtained by filming from a car driving at a steady walking speed of 5 mph*. The camera was held by author PB who was a passenger in the vehicle and who attempted to maintain the focus of expansion in the centre of the camera field. A minimum of 15 minutes of total travel for each speed was collected in up to three sessions. Each movie contained continuous forward motion at a steady speed with minimal acceleration, deceleration or other interruption, from traffic signals or pedestrians. There was no attempt to select the content of the movies for the presence or absence of other vehicles, lamp posts or trees, etc. The gamma function of the camera was calculated by recording a full field movie of a CRT screen containing 16 patches of luminances in linear steps from 0 to 100 cd/m². The corresponding values in the digital movie were fit with a gamma function and the inverse of this function was used to linearise the greyscale of the movies. Movies were captured via a firewire connection and stored to computer disk using Pinnacle Studio software. The resolution was 320×240 at 30 fps and movies were stored as ".avi" files with compression level set to 0. This was half the native resolution of the original movie but retained the full visual field, minimised the file size and thus maximised file reading time.

On each experimental trial, 32 successive movie frames were imported, starting at a random point within the movie each trial, using the MATLABTM function `aviread()`. The red, green and blue values of each frame were gamma corrected and then averaged to produce

*Generating phase noise from the FFT of a noise image, rather than generating random angles directly, ensures the correct (Hermitian) symmetry of the resulting spectrum which in turns ensures a near-zero imaginary result (i.e., that there is no loss of power) when this is back-transformed into the image domain.

a monochrome/greyscale movie of size $320 \times 240 \times 32$ (horizontal pixels \times vertical pixels \times frames). The mean value of the whole movie was subtracted from each pixel, to produce a zero-mean 3D matrix. The first and last 32 columns of each frame were cropped and the first and last eight rows of the movie were padded with zeros to produce a $256 \times 256 \times 32$ movie, centred on the original image.

Masking noise

The amplitude spectrum of the noise was matched to that of the stimulus (and to the sensitivity of the visual system) by randomising the phase-spectrum of the source movie. The 3D amplitude spectrum of a driving movie (calculated by taking the absolute values of the 3D Fast Fourier Transform (FFT) computed with the MATLABTM function `fftn()`) was combined with the phase spectrum (calculated with the MATLABTM function `angle()`) of a $256 \times 256 \times 32$ white noise image. The inverse FFT of this operation produced a random phase noise movie with the same amplitude spectrum as the driving movie.

The random phase noise was notch filtered in the 3D spatio-temporal frequency domain. Notch filters were one octave wide log exponential filters:

$$A(f) = 1 + \left[-1 * \exp\left(\frac{|\ln(f/F_{peak})|^3 \ln 2}{(b_{0.5} \ln 2)^3}\right) \right] \quad (3)$$

where f specifies spatial frequency, temporal frequency or speed, F_{peak} specifies the peak frequency and $b_{0.5}$ the half bandwidth of the filter in octaves, which was fixed at 0.5 octaves (i.e., the full bandwidth was one octave). These filters have the advantage of shallower tails than some alternative band-pass filters (e.g., Laplacian of Gaussian see (Dakin & Bex 2003) for discussion of this property in the context of natural images).

Directional notch filters were linear Gaussians:

$$A(f) = 1 - \left[e^{-\left(\frac{(\theta-F_{peak})^2}{2\sigma^2}\right)} \right] \quad (4)$$

where θ specifies direction; F_{peak} the peak direction of the notch filter, spaced at 22.5 degree intervals from 0 to 360 degrees; and σ specifies the standard deviation, which was fixed at 22.5 degrees.

Before calculating the inverse FFT of the phase randomized masks, the amplitude spectrum of the $256 \times 256 \times 32$ source movie was multiplied by the $256 \times 256 \times 32$ notch filter. The real values of the inverse FFT produced a notch filtered, random phase noise mask.

Contrast scaling

The RMS contrast of the mask movie was fixed at 0.2 (approximately 80% Michelson contrast, although the precise value inevitably varied slightly from movie sample to sample) by multiplying every pixel by 0.2 divided by the standard deviation of all 2097152 pixels. The RMS contrast of the target movie was fixed by a staircase procedure each trial and was scaled in the same way. The sum of the target and mask movies produced a $256 \times 256 \times 32$ movie with zero mean and with the desired RMS contrasts of the target and mask. The target plus mask movie was multiplied by a raised cosine spatio-temporal envelope that was spatially circular with a radius of 8 degrees and a cosine ramp over 16 pixels (0.5 degrees) and a temporal ramp over the 1st and last 3 frames. The movie was scaled between 0–255 by multiplying the 3D matrix by 127 and adding a constant of 127 (note that the RMS contrast of the rescaled image is now calculated as the standard deviation of luminance (or LUT value,

the effect is the same) divided by the mean luminance (or LUT value) and is unchanged by this rescaling). Monochrome resolution was increased to 10.8 bits by bit-stealing (Tyler 1997). We adjusted the luminance of a blue screen (all pixels [0 0 127]) so that it was half that of a red screen ([127 0 0]) which in turn was half that of a green screen ([0 127 0]) at mean luminance, (where white, [127 127 127] was 50 cd/m²). This shifted the white-point of the monitor to 0.31, 0.28 (x,y) at 50 cd/m² (measured with a Minolta CS100 photometer). Under the constraint that no colour could deviate from the others by more than one LUT step, this permitted 1785 unique levels. As a conservative estimate of the precision of our apparatus, the minimum threshold RMS contrast of the images in our experiments was 0.025. Gaussian white noise that is rescaled as in our experiment with this contrast spans the LUT range 113 to 141 (c.11% Michelson contrast). These 28 levels are represented by 196 unique monochrome steps with bit stealing. This resolution is sufficiently high for contrast threshold measurements. Furthermore, this estimate of 28 monochrome levels is conservative because natural images are more leptokurtotic (i.e., the distribution of grey levels has longer tails) than Gaussian noise (Daugman 1988; Field 1994) and so span a greater number of LUT steps for a given RMS contrast.

Each movie frame was updated on alternate video frames of a La Cie Electron 22 blue 21" monitor running at 75 Hz, so that the movie sequence lasted 850 msec. This meant that movies were presented 17% faster than the recording speed (thus 75, 30 and 5 mph movies actually depicted 82, 35 and 6 mph, respectively). The luminance of the monitor was measured with a Minolta CS100 photometer and gamma corrected in the graphics card control panel. We used a GeForce4 MX440 graphics card whose gamma function was linearised directly to produce true linear 8 bit resolution, without any loss of resolution that occurs with look-up-tables. Movies were either played forwards or backwards, at random, across trials. A forward sequence represents expanding optic flow field consistent with forward self motion, the reverse sequence represents contracting optic flow consistent with backwards self motion.

Subjects and procedure

The subjects were two of the authors. Each wore conventional optical correction and viewed the screen monocularly with their dominant eye. Observers fixated a prominent fixation cross at the centre of the screen and were required to report whether the movie simulated forward or reverse motion of the vehicle. Feedback was provided following incorrect responses. Although in real life observers rarely experience high speed backwards self-motion (perhaps only when looking out of the rear window of a vehicle), approaching and receding objects produce expanding or contracting retinal image motion, respectively, over a limited area of the visual field. Even if somewhat unfamiliar, the task was easy in unmasked conditions, and in pilot runs we determined that forward and reverse responses occurred with equal probability.

The RMS contrast of the random phase mask was fixed (at 0.2 or ~80% Michelson contrast), while the RMS contrast of the source movie was under the control of a staircase (Wetherill & Levitt 1965), that reduced the contrast by 1 dB (1/20 log unit) following three correct responses and increased the contrast by 1 dB after one incorrect response. The staircase terminated after ten reversals or 50 trials, whichever occurred first. All conditions for each parameter of the notch filtering (spatial frequency, temporal frequency, speed and direction) were randomly interleaved on a single run. Runs for the four parameters were tested in random order. The raw data from a minimum of five runs for each condition (at least 200 trials per psychometric function) were combined and fit with a cumulative normal

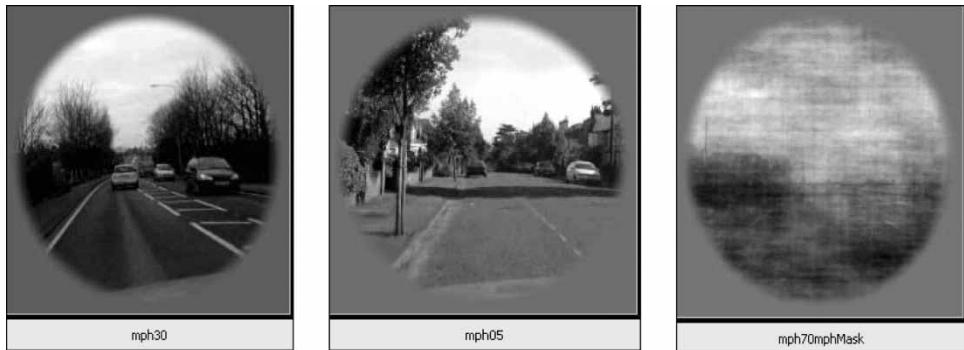


Figure 1. (a) and (b) Examples frames from the original movies. (c) Example sum of movie and random-phase mask.

function by least Chi-square fit (in which the data are weighted by the binomial standard deviation calculated from the observed proportion correct and the number of trials tested at each level). The contrast threshold was estimated from the 75% correct point of the psychometric function and 95% confidence intervals on this point were calculated with a bootstrap procedure, based on 1000 data sets simulated from the number of experimental trials at each level (Foster & Bischof 1991).

Results

Spatial frequency tuning

Our first experiment examined which spatial structure within broadband dynamic scenes observers rely upon to perform a direction discrimination task. In order to quantify the spatial structure of dynamic natural scenes statistically, we first examined the spatial frequency content of a variety of movie images. We calculated the value of α (the slope of log amplitude vs. log spatial frequency in Equation 1) for images captured from various digitized movies, including indoor, outdoor, manmade and natural scenes. The commercial movies were ripped to avi files at native video size (720×576 pixels) and 30 fps with compression set to zero. While these dynamic scenes may not accurately represent the natural environment of many primates, they do constitute on average approximately 15% of the visual diet of television-watching adults (UK Census 2000). The value of α was calculated from the central 256×256 pixels of 100 frames randomly selected from each movie. Red, green and blue pixel values were averaged to produce monochrome image sequences. Amplitudes were averaged within one octave bands and across orientations and were fit with MATLAB's polyfit() function to determine the intersect and slope of log spatial frequency versus log amplitude for each image.

We did not pad the image with zeros to minimise edge artifacts. As a control to determine the effect of padding on the estimate of spectral slope, we generated 1000 noise images with a $1/f^2$ slope. We then estimated the slope of this image as described above with and without zero padding to 512×512 pixels. The slope for unpadded images was on average 1.02 times steeper than that for padded images, but the two were not significantly different ($t = -1.07895$, $p > 0.05$) for this number of samples. Although we measured and corrected the gamma function of our camera, we did not know the gamma function of the camera or processing used to produce commercial movies. As a second control to determine the effect

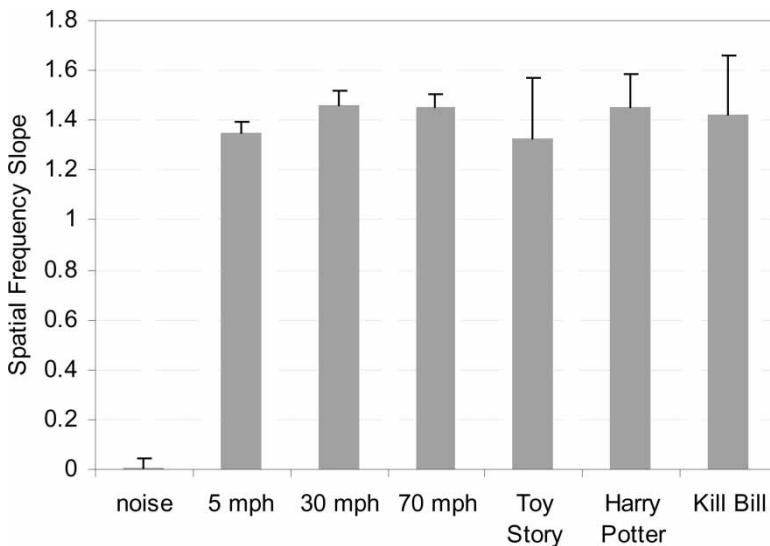


Figure 2. Slope of the log spatial frequency vs. log amplitude spectrum (α) for the movies indicated on the abscissa. Slopes were calculated for 6400 256×256 pixel frames randomly sampled from each movie. Error bars show ± 1 standard deviation. See text for details.

of γ on the estimate of spectral slope, we generated 100 noise images with a $1/f^2$ slope then rescaled the images with gamma values between 0.5 and 4 in log steps. The value of α was then calculated for each rescaled image the value of γ did not significantly affect the value of α ($t < 0.42$ in all cases). We speculate that this is because the changes in the luminance range with γ affect the magnitude of all components approximately equally.

Figure 2 shows the mean value of α for a range of different movie types, including a cartoon and our driving movies at each speed and dynamic noise. The value of α for the movies is consistent with many previous studies of natural images (Billock 1996) and the zero slope for noise images is as expected for white noise. These calculations confirm that the distribution of energy in natural scenes and the log-scaled channels of the visual system are fairly well matched.

The utility of different spatial frequencies components within our movies was measured with log exponential notch filters (Equation 3) centered at 1, 2, 4, 8 or 16 c/deg. The temporal frequency spectrum of the stimuli was otherwise unaltered. Figure 3 shows the RMS contrast of the driving movie that was required to detect its direction of motion (on 75% of trials) when it was embedded in spatial frequency notch filtered noise. At all speeds and for both observers, thresholds were significantly lower for notch filters with a peak at around 2 c/deg (repeated measures ANOVA, $F_{4,8} = 5.77$, $p < 0.05$). Averaged across observers and conditions, thresholds at 2 c/deg are 40% lower than those at 0.5 or 8 c/deg. This means that the direction of motion in noise could be detected at lowest contrast when components at around 2 c/deg were removed from the masking noise. Expressed another way, noise components around 2c/deg had the greatest masking power.

These results are in good agreement with the peak of the contrast sensitivity function for images of this size reported for contrast detection of static (Campbell & Green 1965; Campbell & Robson 1968; Howell & Hess 1978) or drifting sine wave gratings (Kelly 1977; Watson 1986) and with static compound images (Watson 2000; Bex & Makous 2002). However, we

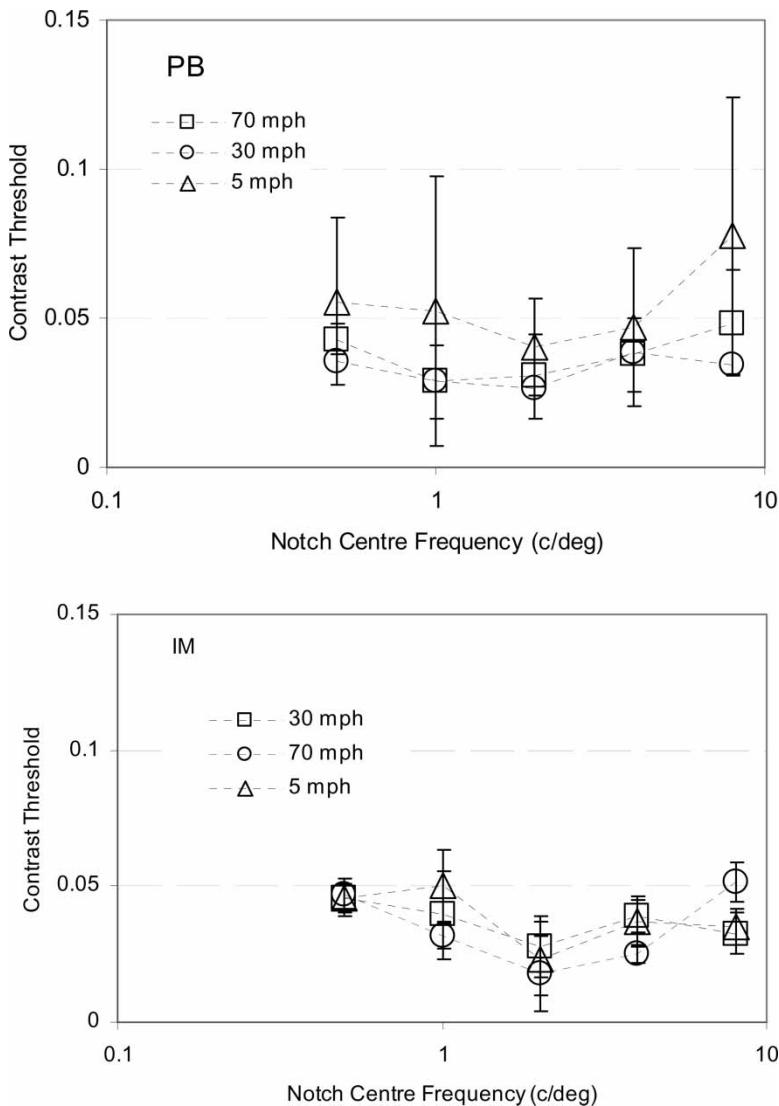


Figure 3. Spatial frequency tuning of critical band masking for optic flow for two observers (PB and IM) at the 3 speeds indicated in the legend. The x-axis shows the peak spatial frequency of the one octave notch that was absent from our spatial frequency filtered noise masks. The y-axis shows the RMS contrast threshold (in RMS units) supporting 75% correct direction discrimination (forwards versus backwards motion). Error bars show 95% confidence intervals.

find that the peak of the function (at around 2c/deg) is the same at all speeds, whereas contrast thresholds for detecting the direction of heading of random dot optic flow patterns has been reported to shift towards low spatial frequencies at higher speeds (Kim & Turano 1999).

Temporal frequency tuning

Our next experiment examined how observers utilise time-varying content within broad-band dynamic scenes. We first examined the temporal frequency content of a variety of

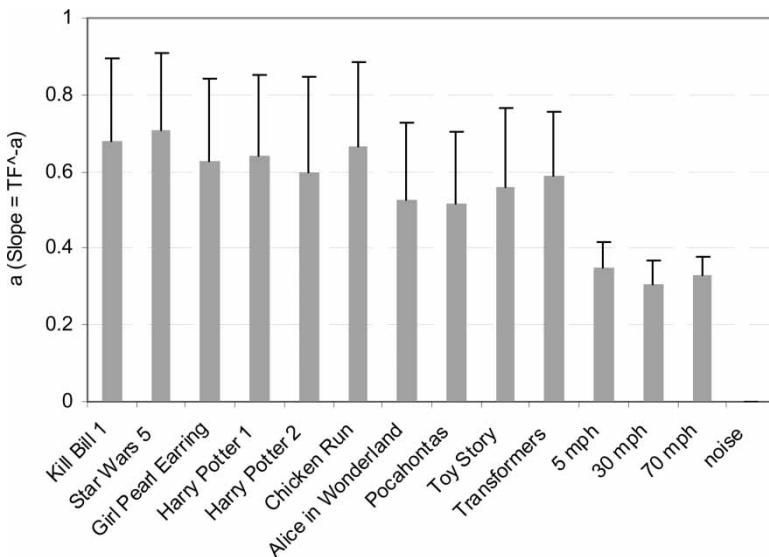


Figure 4. The slope of the log temporal frequency vs. log amplitude spectrum (α) for a range of movies, indicated by the axis. Each estimate of slope is based on 100 movies each 256×256 pixels and lasting 64 frames. Error bars show ± 1 standard deviation. See text for details.

commercial movies by calculating the temporal value of α (the slope of log amplitude vs. log temporal frequency) for 100 randomly selected movie sequences each 256×256 pixels \times 64 sequential frames. The amplitude at each temporal frequency was averaged across all spatial frequencies and orientations. In the commercial movies, we ignored scene cuts in part because we found that scene cuts were surprisingly frequent in 64 frame sequences. Given that human observers make frequent saccadic eye movements that abruptly change the retinal image (Yarbus 1967) and the observation that observers tend to shift fixation to image regions containing different spatial structure (Krieger et al. 2000; Dragoi et al. 2002), we concluded that this simplification was acceptable. Figure 4 shows the mean value of α for a range of different movie types. The slope of white noise is flat, as expected, but amplitude decreased with temporal frequency in dynamic natural scenes. The slope of this attenuation () is consistent with many previous studies (Dong & Atick 1995a; van Hateren 1997; Billock et al. 2001). This is not an artifact of the inclusion of scene cuts in our movie samples—this would tend to reduce the temporal frequency slope (i.e., make it more like temporal white noise, with zero slope), whereas the movie slopes were clearly negative. Furthermore, there are no scene cuts in our driving movies, yet their temporal slopes were shallower than those of the other movies—i.e., they have *relatively* more energy at high temporal frequencies. This is almost certainly because our images contain full field motion across all frames, whereas sequences of 2–3 seconds in commercial movies frequently contain little if any motion. As for the static images in Experiment 1, we estimated the effect of the unknown γ value in the commercial movies on the value of α for temporal frequency. We generated 100 random noise movies of size $256 \times 256 \times 64$ where we fixed the value of α for spatial frequency at 2 and temporal frequency at 1. While had no net effect on the value of α for spatial frequency, the value of α for temporal frequency was progressively underestimated as γ increased. This is to be expected: natural images are dominated by pixels near the mean luminance (Daugman 1988; Field 1994), and as γ increases, this portion of the input-output function is progressively expanded. Thus the effective dynamic range (and thus high

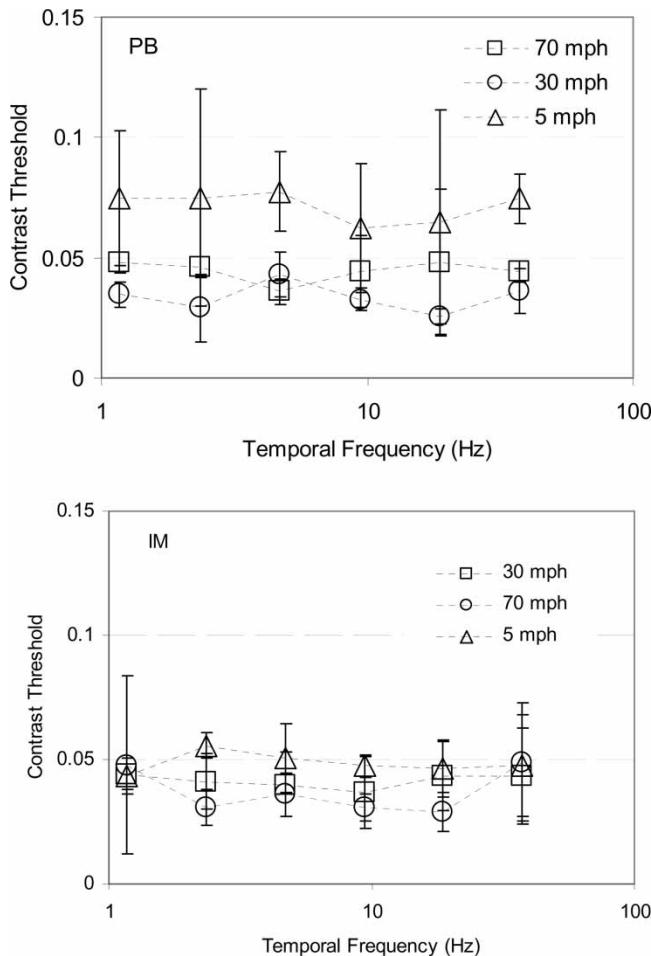


Figure 5. Temporal frequency tuning of critical band masking for optic flow. Details as Figure 3, except that the notch filter removed a one-octave band of temporal (rather than spatial) frequency components.

temporal frequency modulation) of pixels near mean luminance increases with γ . For this reason, the values reported in Figure 4 can therefore only be taken as approximations.

The utility of different temporal frequency components in our movies was measured with log exponential notch temporal frequency filters (Equation 3) centred at 0, 1, 2, 4, 8 or 16 Hz. The spatial frequency spectrum of the stimuli was otherwise unaltered. Figure 5 shows the RMS contrast of the driving movie that was required to detect its direction of motion in temporal notch filtered noise on 75% trials. It is evident that at all speeds and for both observers, the tuning functions were flat. There was no evidence that any temporal frequency was preferred by the visual system to signal the direction of motion in our broadband stimuli. In other words, and contrary to the spatial frequency results, noise at all temporal frequencies has the same masking power.

Speed tuning

The movements of objects in natural scenes are defined by a common speed, rather than a particular spatial or temporal frequency. We wondered whether this property may have

caused the flat tuning functions for temporal frequency. It is possible that the direction discrimination task could have been based on any of the various combinations of spatial and temporal frequency that might have produced some optimally detectable speed. If this were the case, it should be possible to determine such an optimal speed with critical band speed masking.

We first examined the distribution of speeds in a variety of movies, where speed is defined in Equation 2. Many different combinations of spatial and temporal frequency can produce the same speed. The upper panel in Figure 6 shows log amplitude as a function of log speed for 100 random samples of noise and the 30 mph driving movie, each 128×128 pixels and

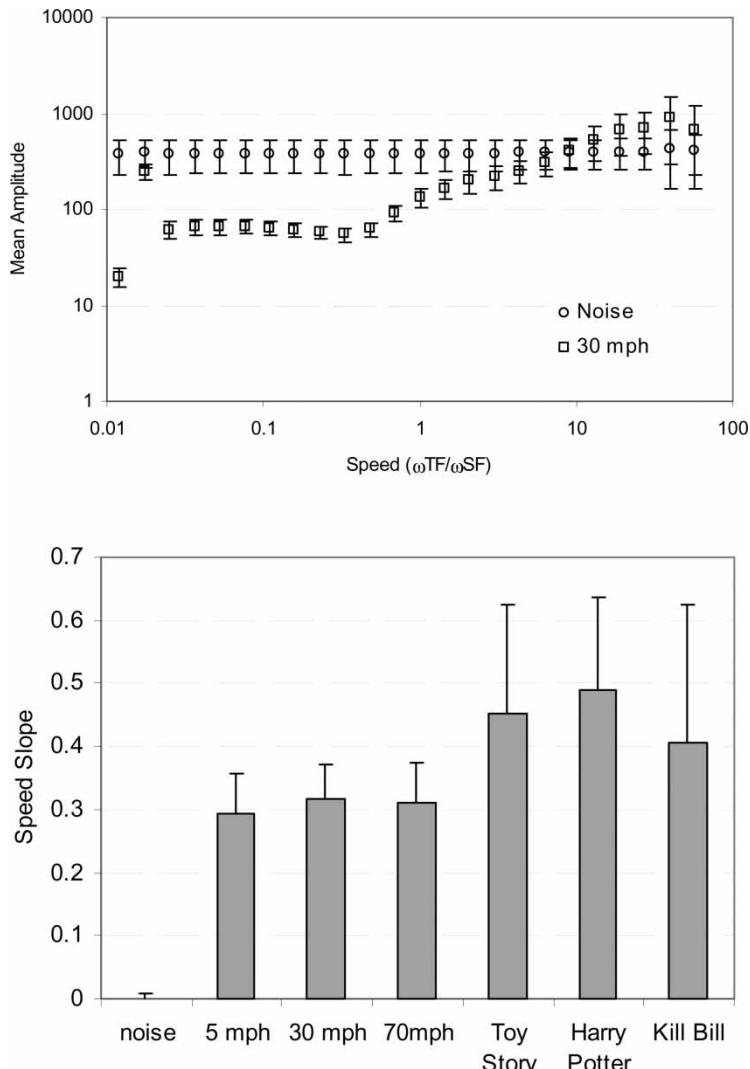


Figure 6. Upper panel: Illustration of log speed vs log amplitude for noise and 30 mph driving movie. The left data point is stationary (zero speed). Lower panel: Slope of the log speed vs log amplitude spectrum for a range of movies, indicated by the axis labels. All data are based on 100 movies each 128×128 pixels* 64 frames. Error bars show 1 standard deviation. See text for details.

64 frames duration. Amplitude was averaged within one octave bands of speed and across all directions/orientations. A straight line was fit to the data on log-log axes which, while failing to capture the elbowed-shape of the curve, does quantify the reduction in amplitude at low speeds. The lower panel in Figure 6 shows the slope of such fits (log amplitude vs. log speed) derived using a random sample of 100 sequences from a range of movies, each of size 128×128 pixels and 64 frames duration. As before, noise has a speed distribution spectrum that is on average flat while real movies have relatively high energy at high speeds. We speculate that the relatively high amplitude at high speeds arises because the rigid motion of edges and contours in real scenes produces modulation at temporal frequency across spatial scales (see Discussion).

The sensitivity of the visual system to different speeds in our movies was measured with log exponential notch speed filters (Equation 3) centered from 1 to 40 deg/sec in log steps. Figure 7 shows the RMS contrast of the driving movie that was required to detect its direction of motion in notch-speed filtered noise on 75% trials. At all speeds and for both observers, the tuning functions were flat. In other words, there was no evidence that any speed was preferred by the visual system to detect direction of motion. Alternatively, all speeds in noise have equal masking power.

Direction tuning

In the final section, we examined how observers' sensitivity to the overall direction of our stimuli depended on the directional content of our movies. We first examined the directional content of 100 randomly selected movie sequences each 256×256 pixels and 64 frames. In order to equate the distribution of components at all orientations/directions, the spatial frequency amplitude spectrum was windowed at all temporal frequencies with a circular annulus whose edges were attenuated with a raised cosine function over 8 pixels. The movie was then filtered with a Gaussian band-pass direction filter with a peak at 22.5 degree intervals and with a standard deviation of 11.25 degrees (see Equation 4). Figure 8 shows the summed amplitude in each direction band for a range of movies. Horizontal and vertical motion clearly dominates most movies, a finding that mirrors earlier work looking at the orientation distribution in static natural images (Switkes et al. 1978; Hancock et al. 1992; van der Schaaf and van Hateren 1996; Coppola et al. 1998; Keil & Cristobal 2000; Hansen et al. 2003; Betsch et al. 2004). Horizontal motion dominates vertical motion in commercial movies; we speculate that this anisotropy is caused by panning camera motion that is frequently employed in the making of commercial films.

The sensitivity of the visual system to different directions of motion in our movies was measured with Gaussian notch orientation filters (Equation 4) centered at 22.5 degree intervals. Figure 9 shows the RMS contrast of the driving movie that was required to detect the direction of motion in notch-speed filtered noise on 75% trials for the 30 mph movies. For both observers, the tuning functions were flat. In other words, there was no evidence that the visual system was more sensitive to, or relied upon, any particular range of directions, even though natural scenes plainly have more energy concentrated along the cardinal axes of motion.

Discussion

Statistical properties of natural scenes

Our results confirm many previous studies that have reported that the amplitude of high spatial frequency components are attenuated in static natural images; the well-known property of “ $1/f$ ” scaling (Kretzmer 1952; Burton & Moorhead 1987; Field 1987; Tolhurst et al. 1992;

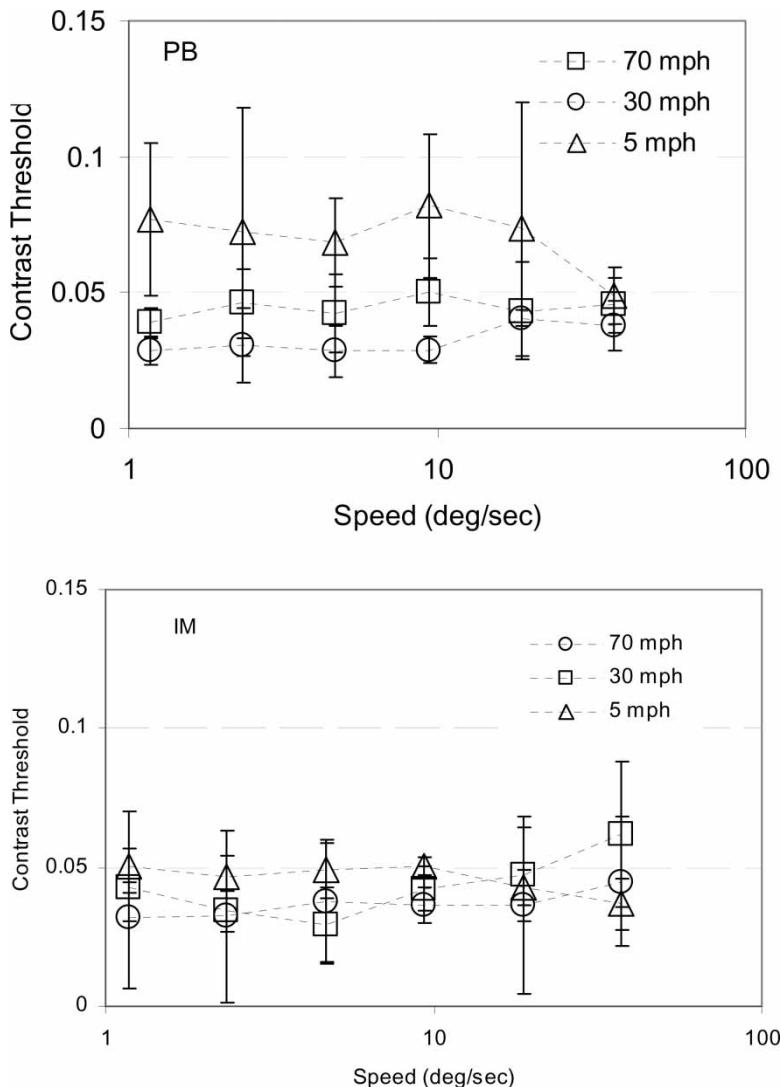


Figure 7. Speed tuning of critical band masking for optic flow. Details as Figure 3, except that the notch filter removed a one-octave band of speed (ω_{TF}/ω_{SF}) components.

Ruderman 1994; Billock 1996; van der Schaaf & van Hateren 1996; van Hateren & van der Schaaf 1998). This property has been attributed to the presence of numerous edges, shadows and occlusions in natural scenes, which have a $1/f$ amplitude spectrum (Tolhurst et al. 1992; Balboa et al. 2001), or to the correlations across space of illuminant and material properties that produce correlations in image luminances (Ruderman 1997). These explanations similarly account for the present results. We also confirm that the amplitude of high temporal frequency components is attenuated in dynamic natural images (Dong & Atick 1995a; van Hateren 1997; Billock et al. 2001). Analogous to the explanations of high spatial-frequency drop-off in static images (Ruderman 1997), this occurs in dynamic natural images because correlations across time of illuminant and material properties produce correlations in image luminances across time.

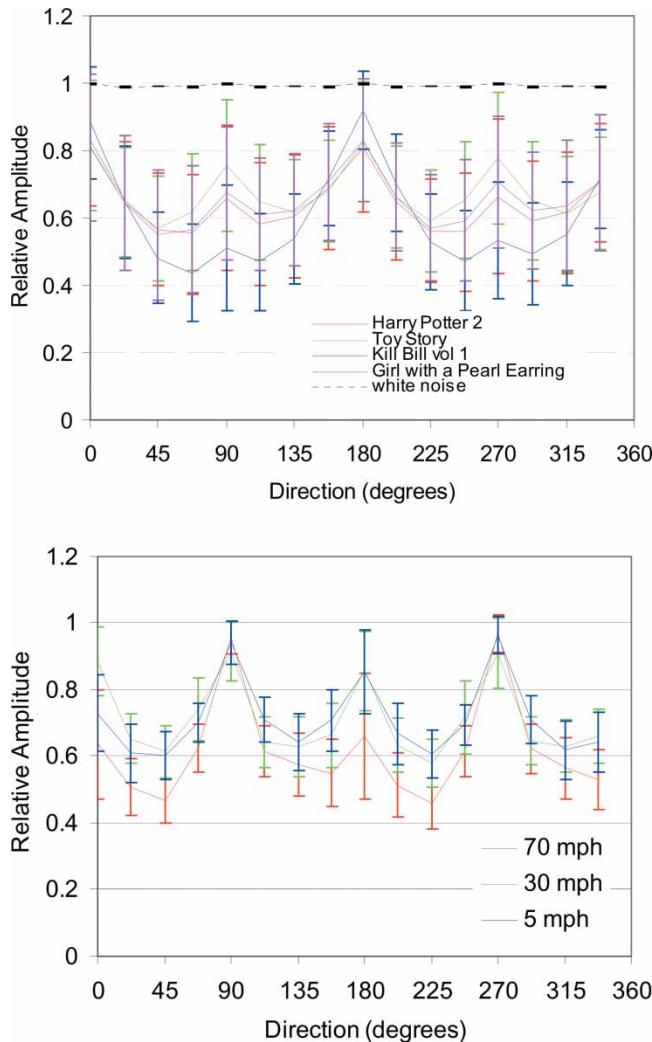


Figure 8. Distribution of direction amplitude for a selection of movies, indicated by the legends. The data show relative amplitude as a function of the direction of motion in the movie. The data for each movie are based on 100 a random movie sequences, each 256×256 pixels $\times 64$ frames duration. Error bars show ± 1 standard deviation. See text for details.

Here, we have extended these analyses of the amplitude spectra of natural images and report that the distribution of speed and direction are also non-uniform in dynamic natural scenes. For dynamic white noise the distribution of speed (averaged across the number of different combinations of spatial and temporal frequency that produce each speed) is flat, like the distribution of spatial and temporal frequency (in fact the flat distribution of spatial and temporal frequency cause this property). However, the total amplitude as a function of speed is not flat and instead reflects the number of combinations of spatial and temporal frequencies that give rise to a particular speed in the FFT. The distribution of speeds in dynamic natural scenes is similarly affected by spatial and temporal frequency combinations in the FFT that produce a given speed, but the mean amplitude across speeds tends to decrease at low speeds. This property seems surprising given that amplitude tends to increase at low spatial and temporal frequencies in dynamic natural images (see Figure 4). At first glance, this would

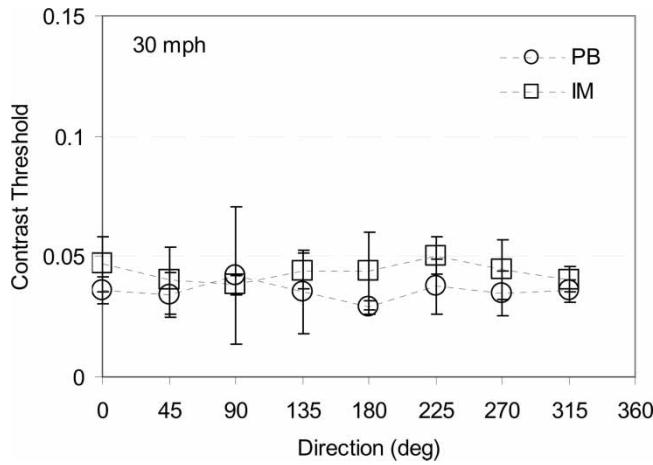


Figure 9. Direction tuning of critical band masking for optic flow. Details as Figure 3, except that tuning was measured for one driving speed (30 mph).

appear to indicate a substantial bias towards low speeds, however, note that there are many more spatio-temporal frequency combinations that produce low speeds in a 3D FFT—mostly by high spatial frequency components that have relatively low amplitude which reduce the mean amplitude at a given speed. Thus, the positive slopes we observe arise because of low energy at low speeds relative to white noise.

We noted above that the slopes of log amplitude vs. log temporal frequency derived from our stimuli were higher than expected for ideal images that are $1/f$ in space time. This suggested to us that the $1/f$ slope for spatial frequency should change with temporal frequency. In order to test this idea, we measured the slope of the amplitude spectrum as in Figure 2, but at each temporal frequency within the 3D amplitude spectrum. Figure 10 shows that for noise, as expected, the amplitude spectrum is flat (zero slope) and this is observed at all temporal

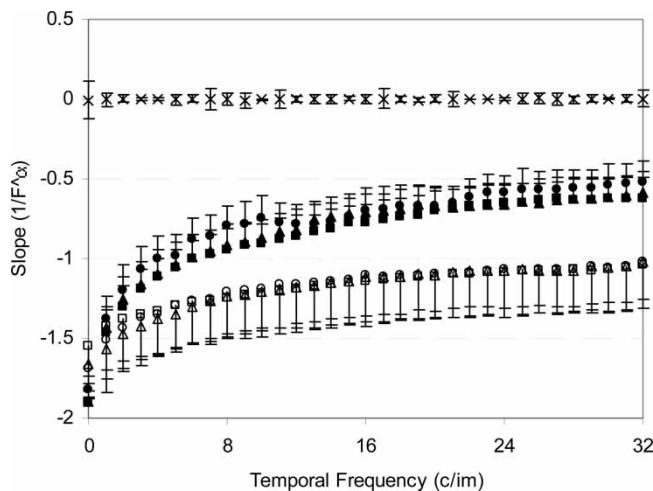


Figure 10. Slope of log amplitude vs. log spatial frequency at different temporal frequencies. The data show the best fitting slope to the log spatial frequency spectrum (as in Figure 2) at each temporal frequency in movies of size 128×128 pixels $\times 64$ frames. Error bars show ± 1 standard deviation for clarity (crosses = noise, filled symbols: circles 5 mph, squares 30 mph, triangles 70 mph; open symbols: circles Kill Bill vol 1, squares Toy Story, triangles Harry Potter and the Philosopher's Stone).

frequencies. However, natural movies show a quite different pattern. The zero frequency (static) component has a slope of around -1.7 , as reported in Results. However, as temporal frequency increases, the slope monotonically decreases (Dong & Atick 1995b). This means that the image structure at high spatial frequencies has a relatively higher amplitude at high temporal frequencies. We speculate that this is caused by the movements of real objects which produce rigid contour motion at a common speed across scales. Temporal frequency must increase with spatial frequency to produce a fixed speed (Equation 2) and so the amplitude of high spatial frequencies must increase relative to that of low spatial frequencies. This is statistically unlikely in dynamic noise, in which spatial frequency structure is uncorrelated across space and time.

We also report that the distribution of motion directions in dynamic natural images is biased toward cardinal axes (horizontal & vertical). This is consistent with a well-known oblique effect in static natural scenes which contain relatively high amplitude at cardinal orientations (Switkes et al. 1978; Hancock et al. 1992; van der Schaaf & van Hateren 1996; Coppola et al. 1998; Keil & Cristobal 2000; Hansen et al. 2003; Betsch et al. 2004). The bias towards vertical motion has been attributed to the presence of the horizon in many natural images; such horizontal contour structure generates predominantly vertical motion energy. The horizontal motion bias can be attributed to the combined influence of the horizontal ground plane and the influence of gravity on the objects that sit upon it—i.e., vertical structure is more stable because gravity exerts no net force. We also show that the commercial movie samples contain a further bias toward horizontal (left/right) motion that is not present in our driving movies. This is consistent with motion along the ground plane of movie objects and actors and with horizontal camera panning sequences in many movies. Our own movies conversely, contain expanding or contracting flow fields with motion in all directions, reducing any of the horizontal/vertical bias.

Spatial frequency tuning

We found a small but significant reduction in contrast direction thresholds when components at around $2\text{--}3$ c/deg were removed from the masking noise suggesting that structure at this spatial frequency is most effective for direction discrimination. The location of this peak in direction sensitivity is consistent with the peak in contrast sensitivity reported for isolated static sinusoidal patterns of this size (Campbell & Green 1965; Campbell & Robson 1968). See Watson (2000) for a review of models. At supra-threshold levels, the location of the peak coincides with the reduction in threshold and supra-threshold apparent contrast that occurs when components at $2\text{--}4$ c/deg are removed or phase randomised in natural images (Bex & Makous 2002). For moving images, this finding is also consistent with the peak spatial frequency (c. 3 c/deg) for the optimal motion stimulus (Watson & Turano 1995).

In a study of sensitivity to the direction of heading with band-pass filtered dot stimuli, Kim and Turano (1999) reported that observers were most sensitive to spatial frequencies at around 2 c/deg and the location of this peak shifted to lower spatial frequencies at higher speeds. The peak observed in the present study is in good agreement with this result, although the peak did not shift with speed. A possible reason that we did not observe a shift is that local temporal frequencies (Figure 4) and speeds (Figure 6) in our stimuli were little affected by our manipulation of driving speed (from 5 to 70 mph). We speculate that this is because under natural conditions, for safety purposes, motorists and road designers increase the distance between vehicles, lanes and boundaries, etc. at high travel speeds to reduce the likelihood of collisions and to increase the time available for compensatory movements. This results in

a relatively invariant distribution of retinal speed that may ensure constant sensitivity over many locomotion speeds, but it undermines our efforts to compare optic flow sensitivity at different natural travel speeds.

Lack of tuning for temporal frequency, speed and direction

Although we observed modest tuning for spatial frequency, there was no change in contrast direction thresholds for any notch filtered temporal frequency, speed or direction filtered noise mask. For isolated grating patterns, peak sensitivity shifts to lower spatial frequencies as temporal frequency increases (Robson 1966; Kelly 1971; Kulikowski & Tolhurst 1973; Tolhurst et al. 1973). It might therefore be possible that observers switch to a different spatial frequency for each temporal frequency notch filter and this somehow results in roughly constant sensitivity. However, the notch filters removed all spatial frequencies at each peak temporal frequency so performance must depend on other temporal frequencies—and here we show that these other temporal frequencies are equally effective.

The peak sensitivity for speed also shifts to lower spatial frequencies as temporal frequency increases, producing roughly constant sensitivity overall (Burr & Ross 1982). This observation is perhaps unsurprising because, when expressed in terms of the local modulation rate of the image, sensitivity is relatively constant. Our notch filters removed different combinations of spatial and temporal frequency and so it is possible that the flat speed tuning we observe in Figure 7 arises because observers switch to different combinations of spatial and temporal frequency under different masking conditions.

It is well established that visual acuity and sensitivity are often worse for stationary stimuli at oblique angles than for identical horizontal or vertical stimuli (for reviews see Appelle (1972) & Essock (1980)). A similar anisotropy favouring cardinal directions has also been observed for detection and direction discrimination thresholds for moving targets (Ball & Sekuler 1982; Heeley & Timney 1988; Heeley et al. 1997; Gros et al. 1998; Matthews & Qian 1999; Loffler & Orbach 2001; Matthews et al. 2001; Westheimer 2003). The fact that the human visual system is more sensitive to horizontal and vertical structure lead us to expect that the direction tuning functions would have peaks at cardinal directions (0° , 90° , 180° and 270°). This expectation was further encouraged by the observation that natural movies have higher amplitude along these axes. Therefore, even if the visual system were not more sensitive to cardinal directions, these directions have higher amplitude in the stimuli and should therefore result in lower contrast detection thresholds. But the tuning functions were flat.

Recently an inverse oblique effect has been reported in the perception of broad-band and natural scenes. In these images, contrast increments on oblique axes are more visible than those on cardinal axes (Essock et al. 2003; Hansen et al. 2003). Hansen et al. (2003) suggest that the higher amplitudes along cardinal axes produce larger responses in neurons tuned to these orientations and this produces greater divisive normalization (Heeger 1992; Wilson & Humanski 1993), which in turn renders an increment discrimination more difficult. An analogous regulation of contrast gain across channels could account for the flat tuning functions observed here.

A failure of random phase masks?

The most frequently used broad-band mask is white noise because pixel values are independent in these images (Pelli & Farell 1999). However, psychophysical estimates of channel

bandwidth (Henning et al. 1981; Losada & Mullen 1995; Solomon 2000) and the tuning of neurons in area V1, the first cortical projection area for visual input, have spatial frequency tuning bandwidths that are approximately log scaled (Campbell et al. 1969; Maffei & Fiorentini 1973; Ikeda & Wright 1975; Movshon et al. 1978; DeValois et al. 1982). This relationship between the tuning of the visual system and the distribution of energy in natural scenes means that the excitation across channels in the primate visual system is relatively uniform for natural images, but is heavily biased towards high spatial frequencies for white noise (Brady & Field 2000). It has previously been argued that this property can explain some otherwise puzzling psychophysical results with moving noise patterns (Bex et al. 1995; Brady et al. 1997; Hess et al. 1998). For example, when only one frame of a random dot kinematogram (RDK) is spatial frequency filtered, motion energy at spatial scales common to both frames should support motion perception, but does not (Morgan & Mather 1994). This finding has been taken as evidence that motion detection is based not on motion energy (Adelson & Bergen 1985; van Santen & Sperling 1985; Watson & Ahumada 1985), but on correspondences between features (such as zero crossings). The locations of such features are uncorrelated between frames in hybrid RDKs and motion detection fails because the visual system cannot solve the correspondence problem (Morgan & Mather 1994). We have argued instead that the visual system is overwhelmed by unpaired noise signals at high spatial scales in the hybrid white noise stimuli. When $1/f^2$ noise (in which energy is distributed evenly across constant log-bandwidth spatial frequency channels) is used instead of white noise, motion perception is possible between all pairs of filtered and unfiltered frames of a RDK. Performance cannot be based on edge matching in these patterns because there is no net correspondence among edges in the two frames. Motion detection must therefore be based on motion signals at spatial scales common to both images (i.e., motion energy).

In the present study, we therefore elected to use a noise mask whose spectrum was matched to that of the stimulus and to the sensitivity of the visual system. However, is it possible that the flat tuning we observed were caused by the random phase masks (with matched amplitude spectra) instead of white noise masks? This is unlikely because, owing to the biases in the amplitude spectra of dynamic natural scenes, if anything our choice should have *overestimated* any tuning. For example, suppose that the visual system were equally sensitive to all spatial frequencies, temporal frequencies and orientations. Any biases in the amplitude spectra of our movies should elicit the largest response in the visual mechanisms that are most sensitive to this component. The most visible stimulus therefore should be a movie that contains the biased components presented with a mask that does not contain them. Since our functions showed no evidence of such tuning it is therefore unlikely that the absence of tuning was caused by an inappropriate choice of mask.

A failure of notch filtered masks?

Previous masking studies have employed band-pass or low-pass and high-pass filtered mask images, with the exception of Losada and Mullen (1995) and Mullen and Losada (1999). Low-pass and high-pass masks have the advantage that no assumptions are necessary about the properties (e.g., bandwidths) of detectors under investigation. However, we favoured *notch filtered* masks, in which all components within a target band are removed from the noise. There are several advantages of such masks:

- (1) All except the target components are masked equally in all conditions. Conversely, low-pass and high-pass filters change the amplitude and bandwidth of masking as the spatial frequency cut-off is changed across conditions.

- (2) Notch filtered masks prevent 'off-frequency looking' (Pelli & Farell 1999), analogous to "off frequency listening" in auditory psychophysics (Patterson 1976). This refers to the use of a channel that is non-optimally tuned for the stimulus in the absence of a mask but which is relatively unaffected by the mask and may provide a misleading estimate of selectivity and tuning.
- (3) Low-pass and high-pass masks allow unspecified pooling across a number of unmasked channels.
- (4) It is easy to equate the RMS contrast of all masking images. This is difficult to achieve with band-pass, low-pass or high-pass filtered noise masks without causing look-up-table overflows when the requested luminance values exceed the range of the graphics card or without severely limiting the range of mask contrasts available by the image with the lowest contrast—typically the high-pass filtered image.
- (5) Compared with low-pass and high-pass filtered masks, notch filtered masks require half the number of trials for the same number of conditions (i.e., a single sweep rather than separate sweeps for low-pass and high-pass masks).

Could the weak tuning we report be attributable to the use of notch filtered masks? This is unlikely because previous studies in auditory (Patterson 1976) and visual (Losada & Mullen 1995; Mullen & Losada 1999) psychophysics have reported similar tuning for notch and band-pass filtered masks. If anything, notch filters produce greater masking effects than band-pass masks along with modest differences in the estimated channel bandwidth (Solomon 2000).

In order to address the concerns about our non-standard (phase random and notch filtered) noise masks, we ran a control experiment to establish whether, with our paradigm, we could replicate other studies that have used low-pass and high-pass filtered white noise. We measured the critical channel for letter identification (Solomon & Pelli 1994; Pelli et al. 2004). RMS contrast thresholds for letter identification were measured in white noise or in random-phase, amplitude-marched masking noise. The 26 letters of the alphabet were generated in lower case Arial font using Adobe Photoshop and saved in uncompressed TIFF format. The letters were negative contrast polarity (dark on light), sized 2 deg (based on the height of letter "x") and centered (at half the maximum horizontal and vertical size) on an 8 deg (256 × 256 pixel) blank background. To generate random phase masks, the 26 letters were read from image files into MATLAB and scaled to zero mean and constant RMS contrast. The scaled letters were summed to produce an "average letter". Each trial, a new mask was created by randomising the phase spectrum of the average letter (as before, by combining the amplitude spectrum of the average letter with the phase spectrum of a new random noise sample). The pixel values were then scaled to fix the RMS contrast of the mask at 0.2 each trial. For comparison, white noise masks were generated with the MATLAB function `rand()` and were also scaled to zero mean and 0.2 RMS contrast. The noise masks were then notch filtered with log exponential filters (Equation 3) with peaks varying from 0.125 to 16 cycles per degree (corresponding to 0.06 to 7.3 cycles per letter) in log steps.

On each trial a target letter was selected at random and scaled to have zero mean and an RMS contrast specified by a 3-down, 1-up staircase. The letter and mask were summed and scaled between 0–255 with 10.8 bit resolution provided by bit-stealing (Tyler 1997) within a circular raised cosine window subtending 8 deg at the 57 cm viewing distance. The stimuli were presented for 500 ms with abrupt onset and offset. The observer's task was to identify the target letter by clicking the cursor on the corresponding letter presented in two rows along the top of the screen. Feedback was provided following incorrect responses. Figure 11 plots RMS contrast thresholds for 26 AFC letter identification as a function of the

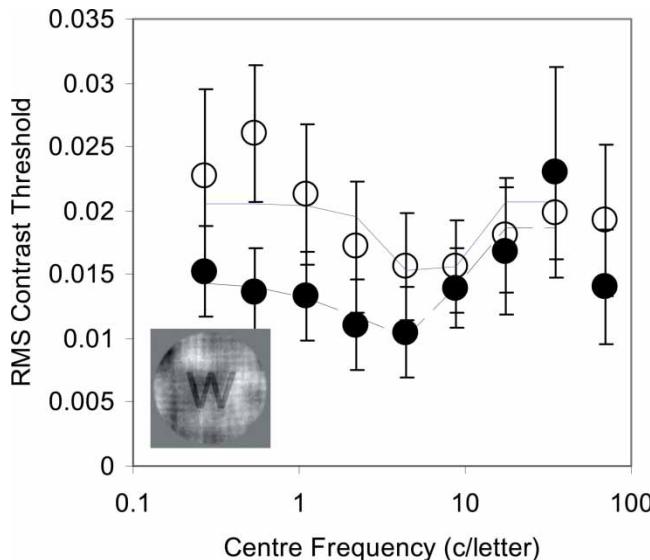


Figure 11. Critical band masking for stationary letter identification. As Figure 3, except that the target was 1 of 26 Arial font letters presented in white noise masks (filled symbols) or random phase average letter masks (see text for details). Separated data on the right are for unfiltered masks. Inset image illustrates typical trial stimulus, in this case a letter W in random-phase mask. Error bars show 95% confidence intervals. Curves are Gaussian fits to the data.

centre frequency of the notch filter. Filled symbols show thresholds for white noise masks; open symbols for random phase masks. Letter identification is band-pass tuned for spatial frequency with both masks. Thresholds are lowest when components at 4.6 cycles per letter are removed from white noise masks, in good agreement with previous estimates of peak sensitivity for letters identification (Solomon & Pelli 1994)[‡]. With average letter random-phase masks, the function is clearly band-pass tuned although the peak is shifted toward a slightly higher spatial frequency (6.5 c/letter). This is consistent with the relatively high amplitude of high spatial frequencies in white noise, compared to the average letter, whose amplitude spectrum (calculated as described above for natural images) falls as $1/f^{1.8}$ (the slopes of the 26 individual letters measured individually have a mean of 1.8 and a standard deviation of 0.15). The key feature of these results is that band-pass tuning is observed clearly with random phase, notch filtered masks.

The present results cannot therefore be attributed to the masks employed in the present study but instead show that the visual system relies upon all spatial and temporal structure derived from real dynamic scenes, approximately equally to perform our direction of heading task.

Conclusions

It has been suggested that the sensitivity of the visual system is influenced by the natural environment in which humans have evolved or developed. Several investigators have proposed that the visual system may be adapted to exploit the statistical redundancies in natural scenes and in so doing maximise the efficiency with which natural images are encoded (Attneave 1954; Barlow 1961; Laughlin 1981; Srivinansan et al. 1982; Field 1987; Van Hateren 1992;

van Hateren & van der Schaaf 1998; Parraga et al. 2000; Thomson et al. 2000; Tolhurst & Tadmor 2000).

We confirm and extend previous studies that have reported statistical regularities in many different natural images. We report that for a direction discrimination task, observers are equally sensitive to components at all temporal frequencies, speeds and directions in high contrast, dynamic natural images. Our results support an emerging view that at supra-threshold contrasts, the responses of the visual system are relatively invariant of spatial scale in natural images and that functional behaviour, such as estimation of direction of heading, depends more on the relationships among image components than on their contrast detection threshold. At supra-threshold contrasts, in spite of large biases in the distribution of the amplitude spectra of natural images, the responses of the visual system are relatively constant and render the components of natural images equally visible to the visual system. Under this view, structure at any spatial frequency, temporal frequency, speed or direction is equally effective for visually-guided behaviour.

Acknowledgments

This research was supported by The Wellcome Trust and by the Biological and Biotechnical Sciences Research Council.

References

Adelson EH, Bergen JR. 1985. Spatiotemporal energy models for the perception of motion. *J Opt Soc America A* 2:284–299.

Appelle S. 1972. Perception and discrimination as a function of stimulus orientation: the oblique effect in man and animals. *Psych Bull* 78:266–278.

Attneave F. 1954. Some informational aspects of visual perception. *Psychol Rev* 61:183–193.

Balboa RM, Tyler CW, Grzywacz NM. 2001. Occlusions contribute to scaling in natural images. *Vision Res* 41:955–964.

Ball K, Sekuler R, Machamer J. 1983. Direction and identification of moving targets. *Vision Res* 23:229–238.

Barlow HB. 1961. Possible principles underlying the transformation of sensory messages. In Rosenblith WA, editor. *Sensory communication*. Cambridge, MA: MIT Press. pp 217–234.

Betsch BY, Einhäuser W, Kording KP, Konig P. 2004. The world from a cat's perspective—statistics of natural videos. *Bio Cybernetics* 90:41–50.

Bex PJ, Brady N, Fredericksen RE, Hess RF. 1995. Energetic motion detection. *Nature* 378:670–672.

Bex PJ, Makous W. 2002. Spatial frequency, phase, and the contrast of natural images. *J Opt Soc America A* 19:1096–1106.

Billock VA. 1996. Fractal properties of natural images and spatial vision. *Investigative Ophthalmology Visual Sci*, 37:4209–4209.

Billock VA, de Guzman GC, Kelso JAS. 2001. Fractal time and $1/f$ spectra in dynamic images and human vision. *Physica D-Nonlinear Phenomena* 148:136–146.

Blakemore C, Campbell FW. 1969. On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *J Physiology* 203:237–260.

Blakemore C, Muncey JPJ, Ridley RM. 1973. Stimulus specificity in the human visual system. *Vision Res* 13:1915–1931.

Bowker DO. 1983. Suprathreshold spatiotemporal response characteristics of the human visual system. *J Opt Soc America* 73:436–440.

Brady N, Bex PJ, Fredericksen RE. 1997. Independent coding across spatial scales in moving fractal images. *Vision Res* 37:1873–1883.

Brady N, Field DJ. 1995. What's constant in contrast constancy? The effects of scaling on the perceived contrast of bandpass patterns. *Vision Res* 35:739–756.

Brady N, Field DJ. 2000. Local contrast in natural images: normalisation and coding efficiency. *Perception* 29:1041–1055.

Bryngdahl O. 1966. Characteristics of the visual system: psychophysical measurements of the response to spatial sine-wave stimuli in the photopic region. *J Opt Soc America* 56:811–821.

Burr DC, Ross J. 1982. Contrast sensitivity at high velocities. *Vision Res* 22:479–484.

Burton GJ, Moorhead IR. 1987. Color and spatial structure in natural scenes. *Applied Optics* 26:157–170.

Campbell FW, Cooper GF, Enroth-Cugell C. 1969. The spatial selectivity of visual cells of the cat. *J. Physiology* 203:223–235.

Campbell FW, Green DG. 1965. Optical and retinal factors affecting visual resolution. *J Physiology* 181:576–593.

Campbell FW, Robson JG. 1968. Application of Fourier analysis to the visibility of gratings. *J Physiology* 197:551–566.

Cannon MW. 1979. Contrast sensation: a linear function of stimulus contrast. *Vision Res* 19:1045–1052.

Coppola DM, Purves HR, McCoy AN, Purves D. 1998. The distribution of oriented contours in the real world. *Proc Nat Acad Sciences USA* 95:4002–4006.

Dakin SC, Bex PJ. 2003. Natural image statistics mediate brightness “filling in”. *Proc Royal Soc London: Series B-Biological Sci* 270:2341–2348.

Daugman JG. 1988. Complete discrete 2D gabor Transforms by neural networks for image analysis and compression. *IEEE Trans on ASSP* 36:1169–1179.

DeValois RL, Albrecht DC, Thorell LG. 1982. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res* 22:545–559.

Dong DW, Atick JJ. 1995a. Statistics of natural time-varying images. *Network: Comp Neural Systems* 6:345–358.

Dong DW, Atick JJ. 1995b. Temporal decorrelation—A theory of lagged and nonlagged responses in the lateral geniculate-nucleus. *Network: Comp Neural Systems* 6:159–178.

Dragoi V, Sharma J, Miller EK, Sur M. 2002. Dynamics of neuronal sensitivity in visual cortex and local feature discrimination. *Nature Neuroscience* 5:883–891.

Essock EA. 1980. The oblique effect for stimulus identification considered with respect to two classes of oblique effects. *Perception* 9:37–46.

Essock EA, DeFord JK, Hansen BC, Sinai MJ. 2003. Oblique stimuli are seen best not worst! broad-band stimuli: a horizontal effect. *Vision Res* 43:1329–1335.

Field DJ. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc America A* 4:2379–2394.

Field DJ. 1994. What is the goal of sensory coding? *Neural Comp* 6:559–601.

Foster DH, Bischof WF. 1991. Thresholds from psychometric functions: superiority of bootstrap to incremental and probit variance estimators. *Psychological Bull* 109:152–159.

Georges MA, Sullivan GD. 1975. Contrast constancy: deblurring in human vision by spatial frequency channels. *J Physiology* 252:627–656.

Gibson JJ. 1966. The senses considered as perceptual systems. Boston: Houghton Mifflin.

Gold J, Bennett PJ, Sekuler AB. 1999. Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Res* 39:3537–3560.

Graham N, Nachmias J. 1971. Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple channel models. *Vision Res* 11:251–259.

Gros BL, Blake R, Hirsh E. 1998. Anisotropies in visual motion perception: a fresh look. *J Opt Soc America A* 15:2003–2011.

Hancock PJB, Baddeley RJ, Smith LS. 1992. The principal components of natural images. *Network: Comp Neural Systems* 3:61–70.

Hansen BC, Essock EA, Zheng WF, DeFord JK. 2003. Perceptual anisotropies in visual processing and their relation to natural image statistics. *Network: Comp Neural Systems*, 14:501–526.

Heeger DJ. 1992. Normalization of cell responses in cat striate cortex. *Visual Neuroscience* 9:181–197.

Heeley DW, Buchanan-Smith HM, Cromwell JA, Wright JS. 1997. The oblique effect in orientation acuity. *Vision Res* 37:235–242.

Heeley DW, Timney B. 1988. Meridional anisotropies of orientation discrimination for sine wave gratings. *Vision Res* 28:337–344.

Henning GB, Hertz BG, Hinton JL. 1981. Effects of different hypothetical detection mechanisms on the shape of spatial-frequency filters inferred from masking experiments: I. Noise masks. *J Opt Soc America* 71:574–581.

Hess RF, Bex PJ, Fredericksen ER, Brady N. 1998. Is human motion detection subserved by a single or multiple channel mechanism? *Vision Res* 38:259–266.

Howell ER, Hess RF. 1978. The functional area for summation to threshold for sinusoidal gratings. *Vision Res* 18:369–374.

Ikeda H, Wright MJ. 1975. Spatial and temporal properties of “sustained” and “transient” neurones in area 17 of the cat’s visual cortex. *Exp Brain Res* 22:363–383.

Keil MS, Cristobal G. 2000. Separating the chaff from the wheat: possible origins of the oblique effect. *J Opt Soc America A* 17:697–710.

Kelly DH. 1971. Theory of flicker and transient responses. II. Counterphase gratings. *J Opt Soc America* 61:632–640.

Kelly DH. 1977. Visual contrast sensitivity. *Optica Acta* 24:107–129.

Kim J, Turano KA. 1999. Optimal spatial frequencies for discrimination of motion direction in optic flow patterns. *Vision Res* 39:3175–3185.

Koenderink JJ. 1986. Optic flow. *Vision Res* 26:161–179.

Kretzmer ER. 1952. Statistics of television signals. *Bell Systems Technical J* 31:751–763.

Krieger G, Rentschler I, Hauske G, Schill K, Zetzsche C. 2000. Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. *Spatial Vision* 13:201–214.

Kulikowski JJ. 1976. Effective contrast constancy and linearity of contrast sensation. *Vision Res* 16:1419–1431.

Kulikowski JJ, Tolhurst DJ. 1973. Psychophysical evidence for sustained and transient detectors in human vision. *J Physiology* 232:149–162.

Laughlin SB. 1981. A simple coding procedure enhances a neuron's information capacity. *Zeitschrift für Naturforschung, Section C, Biosciences* 36:910–912.

Loffler G, Orbach HS. 2001. Anisotropy in judging the absolute direction of motion. *Vision Res* 41:3677–3692.

Losada MA, Mullen KT. 1995. Color and luminance spatial tuning estimated by noise masking in the absence of off-frequency looking. *J Opt Soc America A* 12:250–260.

Maffei L, Fiorentini A. 1973. The visual cortex as a spatial frequency analyser. *Vision Res* 13:1255–1267.

Matthews N, Liu ZL, Qian N. 2001. The effect of orientation learning on contrast sensitivity. *Vision Res* 41:463–471.

Matthews N, Qian N. 1999. Axis-of-motion affects direction discrimination, not speed discrimination. *Vision Res* 39:2205–2211.

Morgan MJ, Mather G. 1994. Motion discrimination in two-frame sequences with differing spatial frequency content. *Vision Res* 34:197–208.

Movshon JA, Thompson ID, Tolhurst DJ. 1978. Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *J Physiology* 283:101–120.

Mullen KT, Losada MA. 1999. The spatial tuning of color and luminance peripheral vision measured with notch filtered noise masking. *Vision Res* 39:721–731.

Parraga CA, Troscianko T, Tolhurst DJ. 2000. The human visual system is optimised for processing the spatial information in natural visual images. *Current Biology* 10:35–38.

Patterson RD. 1976. Auditory filter shapes derived with noise stimuli. *J Acoustical Soc America* 59:640–654.

Peli E, Lee E, Trempe CL, Buzney S. 1994. Image-enhancement for the visually-impaired—the effects of enhancement on face recognition. *J Opt Soc America* 11:1929–1939.

Pelli DG, Farell B. 1999. Why use noise? *J Opt Soc America A* 16:647–653.

Pelli DG, Levi DM, Chung STL. 2004. Using visual noise to characterize amblyopic letter identification. *J Vision* 4:904–920.

Robson JG. 1966. Spatial and temporal contrast sensitivity functions of the human visual system. *J Opt Soc America* 56:1141–1142.

Ruderman DL. 1997. Origins of scaling in natural images. *Vision Res* 37:3385–3398.

Ruderman DL. 1994. The statistics of natural images. *Network: Comp Neural Systems* 5:517–548.

Solomon JA. 2000. Channel selection with non-white-noise masks. *J Opt Soc America A* 17:986–993.

Solomon JA, Pelli DG. 1994. The visual filter mediating letter identification. *Nature* 369:395–397.

Srinivasan MV, Laughlin SB, Dubs A. 1982. Predictive coding: a fresh view of inhibition in the retina. *Proc Royal Soc London B* 216:427–459.

St John R, Timney B, Armstrong KE, Szpak AB. 1987. Changes in perceived contrast of suprathreshold gratings as a function of orientation and spatial frequency. *Spatial Vision* 2:223–232.

Switkes E, Mayer MJ, Sloan JA. 1978. Spatial frequency analysis of the visual environment: anisotropy and the carpentered environment hypothesis. *Vision Res* 18:1393–1399.

Thomson MGA, Foster DH, Summers RJ. 2000. Human sensitivity to phase perturbations in natural images: a statistical framework. *Perception* 29:1057–1069.

Tolhurst DJ, Sharpe CR, Hart G. 1973. The analysis of the drift rate of moving sinusoidal gratings. *Vision Res* 13:2545–2555.

Tolhurst DJ, Tadmor Y. 2000. Discrimination of spectrally blended natural images: optimisation of the human visual system for encoding natural images. *Perception* 29:1087–1100.

Tolhurst DJ, Tadmor Y, Chao T. 1992. Amplitude spectra of natural images. *Ophthalmic Physiol Opt* 12:229–232.

Tolhurst DJ, Thompson ID. 1981. On the variety of spatial frequency selectivities shown by neurons in area 17 of the cat. *Proc Royal Soc London B* 213:183–199.

Tyler CW. 1997. Colour bit-stealing to enhance the luminance resolution of digital displays on a single pixel basis. *Spatial Vision* 10:369–377.

van der Schaaf A, van Hateren JH. 1996. Modelling the power spectra of natural images: statistics and information. *Vision Res* 36:2759–2770.

van Hateren JH. 1992. Real and optimal neural images in early vision. 360:68–70.

van Hateren JH, van der Schaaf A. 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc Royal Soc London B* 265:359–366.

van Hateren JH. 1997. Processing of natural time series of intensities by the visual system of the blowfly. *Vision Res* 37:3407–3416.

van Santen JP, Sperling G. 1985. Elaborated Reichardt detectors. *J Opt Soc America A* 2:300–321.

Watanabe A, Mori T, Nagata S, Hiwatashi K. 1968. Spatial sine wave responses of the human visual system. *Vision Res* 8:1245–1263.

Watson AB. 1986. Temporal sensitivity. In: Boff K, Kaufman L, Thomas J, editors. *Handbook of perception and human performance*, 1. New York: Wiley. pp 1–43.

Watson AB. 2000. Visual detection of spatial contrast patterns: evaluation of five simple models. *Optics Express* 6:12–33.

Watson AB, Ahumada AJ Jr. 1985. Model of human visual-motion sensing. *J Opt Soc America A* 2:322–342.

Watson AB, Turano K. 1995. The optimal motion stimulus. *Vision Res* 35:325–336.

Westheimer G. 2003. Meridional anisotropy in visual processing: implications for the neural site of the oblique effect. *Vision Res* 43:2281–2289.

Wetherill GB, Levitt H. 1965. Sequential estimation of points on a psychometric function. *Brit J Math Stat Psychol* 18, 1–10.

Wilson HR, Humanski R. 1993. Spatial frequency adaptation and contrast gain control. *Vision Res* 33:1133–1149.

Yarbus AL. 1967. *Eye movements and vision*. New York: Plenum Press.