The perception of suprathreshold contrast and fast adaptive filtering

Peter J. Bex  
Schepens Eye Research Institute, Harvard Medical School, Boston, MA, USA

Keith Langley  
Department of Psychology, University College London, London, UK

We examine how the perceived contrast of dynamic noise images depends upon temporal frequency (TF) and mean luminance. A novel stepwise suprathreshold matching paradigm shows that both threshold and suprathreshold contrast sensitivity functions may be described by an inverted-U shape as a function of TF. The shape and the peak TF of the tuning function vary with the conditions under which it is measured. Spatiotemporal vision is weakly band-pass at low luminance levels (0.8 cd/m²) but becomes more strongly band-pass at high luminances (40–400 cd/m²). The peak temporal frequencies of the band-pass functions increase with the mean luminance and contrast of the test signals. As a function of increasing image contrast, our results demonstrate that the visual system broadens the spatiotemporal bandwidth of its signal detection mechanisms, especially at high mean luminances. Our results are shown to be consistent with an adaptable signal transmission system in which early luminance-dependent gain control mechanisms, in combination with on-line estimates of contrast via the autocorrelation function lead to an adaptive enhancement of spatiotemporal vision at high temporal frequencies.

Keywords: temporal vision, contrast, constancy, gain control, autocorrelation function, noise


Introduction

Human sensitivity to sinusoidal modulations in luminance has a classic inverted-U shape, peaking at around 2–4 c/deg (Campbell & Green, 1965; Campbell & Robson, 1968), that reflects the underlying sensitivity of a set of log-spaced spatial frequency (SF) selective channels (Blakemore & Campbell, 1969; Campbell & Robson, 1968; Graham & Nachmias, 1971). At suprathreshold signal levels, suprathreshold apparent contrast (STAC) is relatively invariant of SF (Blakemore, Muncey, & Ridley, 1973; Bowker, 1983; Brady & Field, 1995; Bryndahl, 1966; Cannon, 1979; Georgeson & Sullivan, 1975; Kulikowski, 1976; St. John, Timney, Armstrong, & Szpak, 1987; Watanabe, Mori, Nagata, & Hiwatashi, 1968), a phenomenon termed “contrast constancy” (Georgeson & Sullivan, 1975; for a review, see Georgeson, 1990).

Contrast sensitivity also varies with the temporal frequency (TF) of the stimulus, declining sharply above 5–10 Hz and showing a low frequency decline that depends on SF (Bowker, 1983; Kelly, 1975; Koenderink & van Doorn, 1979; Robson, 1966). At suprathreshold contrasts, constancy is reported to flatten the STAC response function as a function of TF (Bowker, 1983; Camisa & Zemon, 1985; Georgeson, 1987), suggesting that contrast constancy may be equally applied to TF- as well as SF-dependent signals.

There are at least three ways in which contrast constancy could arise in the visual system. The first is where the visual system employs several spatiotemporal tuned channels whose gains are adjusted in such a way that contrast constancy occurs: for example, through an enhancement of the high and low or the suppression of the mid SF- or TF-tuned channels (Georgeson & Sullivan, 1975; Swanson, Georgeson, & Wilson, 1988; Swanson, Wilson, & Giese, 1984). These gain control processes must be rapid as constancy is observed under the short presentation times (~1–2 s) that are used to explore contrast constancy effects.

A second account of contrast constancy, which does not rely upon variable gain control mechanisms, has been proposed by Brady and Field (1995) and by Cannon (1979). They argued that practical difficulties in making matches near detection threshold and the inclusion of zero contrast matches may have obscured what are actually linear contrast response functions. In supposing response functions are approximately linear as a function of contrast and invariant of SF, Brady and Field have also suggested that contrast matching curves could become relatively flatter for high compared with low contrast signals because of an increase in underlying signal-to-noise ratios (Brady & Field, 1995; Cannon, 1979; Kulikowski, 1976).

The explanation for contrast constancy offered by Brady and Field (1995) depends upon the statistics of the
A signal transmission model of contrast constancy

In considering the possibility that the phenomenon of contrast constancy involves precortical mechanisms, we refer to a signal transmission system illustrated in Figure 1. The transmission system is intended to model the propagation of spatiotemporal information from one (neural) location to another. The system possesses three key elements: (i) signal encoding, (ii) signal transmission, and (iii) signal recovery or extraction. The aim of the transmission system is to minimize the MSE between the input and the output signals subject to a constraint placed upon the propagation of information across the communication channel. The propagation constraint can be viewed as analogous to the limited dynamic range attributed to real neurons in the visual system (Laughlin, 1994): a bandwidth constraint. Note that the bandwidth constraint is a critical component of the signal transmission system. In its absence, an optimized signal transmission system would simply swamp the channel noise source with amplified signal at the stage of signal encoding (Diamantaras, Hornik, & Strintzis, 1999; Wainwright, 1999). In many applications, this amplification process would be impractical for a physically realizable system because signal power is often held at a premium.

The model for an optimum signal transmission system is fully explicated in Appendices A and B. In the development of the model, it is assumed without loss in generality that the spatiotemporal computations are local, that is, may be implemented in a neural system using a local window of spatial support. The model, which is derived from this theory, has three stages:

1. Gain control \(G(\mu)\): The model proposes that the visual signal is first suppressed by early luminance gain.

2. Linear compression (Encoder):

3. Linear decompression (Decoder):

![Figure 1](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933520/ on 10/25/2018)
control, whose purpose is to maintain a relatively constant mean output (independent of the actual mean luminance of the signal). The neural site of the gain control mechanism, by virtue of its luminance dependency, is probably retinal (Heinrich & Bach, 2001; Laughlin, 1994; Mante et al., 2005; Rushton, 1965; Snippe, Poot, & van Hateren, 2000; Webster, Georgeson, & Webster, 2002; Wilson, 1997). As can be observed from Figure 1, an early gain control mechanism that post-cedes a source of input noise will attenuate both signal and input noise, thus preserving signal-to-noise ratios. The system’s response (output) at this stage, although constant by design, will nonetheless deliver a signal-to-noise ratio that increases in proportion with increases in signal by virtue of the attenuation of the input noise variance.

(2) A temporal encoding filter, $E(\omega)$, whose TF tuning is determined by the statistical properties of the transmitted and intrinsic noise signals and the transmission channel constraint. The signal’s statistics are represented by the temporal autocorrelation function (TAF), which for this paper is a function of the expected mean, variance, and bandwidth of visual signals (Equation D1). The TAF may also be thought of as an environmental model (Barlow, 1961; Dakin & Bex, 2003; Foldiak & Barlow, 1989; Langley, 2005; Simoncelli, 2003; van Hateren, 1997). The TAF employed here was first derived by Franks (1969) as a mathematical model for video signals (Appendix D). This TAF is given by

$$R_{tt}(\omega, \sigma_{tt}, A, \mu) = \frac{\sigma_{tt}^2}{A} \left[ 1 + \frac{\omega^2}{A^2} \right] + \mu^2 \delta(0), \quad (1)$$

where $\sigma_{tt}^2$ and $\mu$ represent estimates of the spatiotemporal signal’s variance and mean luminance, respectively. $A$ is a rate parameter that controls the bandwidth of the TAF. Smaller values for $A$ imply that visual signals are expected to be more correlated as compared with higher values. In its discrete form (i.e., by taking $Z$ transforms), it is not difficult to show that the parameters depicted in Equation 1 provide a simple characterization of spatiotemporal signals in terms of the expected mean ($\mu$), variance ($\sigma_{tt}^2$), and temporal instantaneous frequency ($\omega(A)$). Numerical values for each parameter may be estimated by a system on-line through separate computations in parallel: the system identification problem (Gelb, 1974; Langley, 2005). Numerical values for the variance ($\sigma_{tt}^2$) linearly increased with stimulus contrast under experimental manipulation, while necessary values for the rate parameter ($A$) and mean ($\mu$) were estimated from the data.

As shown in Appendix B, the temporal tuning of the TAF in Equation 1 is low-pass. Its amplitude spectrum falls approximately as “$1/\omega^5$” for $\omega \gg A$, which is a characteristic of the temporal amplitude spectrum dynamic natural images (Atick, 1992; Bex, Dakin, & Mareschal, 2005; Billock, de Guzman, & Scott Kelso, 2001; Dong & Atick, 1995; van Hateren, 1997).

In assuming that the visual encoding filter may be derived from a class of optimal functions, the temporal encoding filter, $E(\omega)$, that delivers the smallest $MSE$ subject to a transmission constraint (Appendix A; Equation A5) is given by

$$[E^2(\omega)]^+ = \left( \frac{R_{tt}(\omega)}{R_{tt}(\omega) + G^2(\mu)} \right) \sqrt{ \frac{\sigma_{\text{ch}}^2}{\lambda_E R_{tt}(\omega) + G^2(\mu) \sigma_{\text{in}}^2} } - \frac{\sigma_{\text{ch}}^2}{R_{tt}(\omega) + G^2(\mu) \sigma_{\text{in}}^2}, \quad (2)$$

where the notation $[,]^+$ denotes the positive part and $\sigma_{\text{in}}$ and $\sigma_{\text{ch}}$ are the standard deviations of the input and channel noise signals, respectively (Franks, 1969). $\lambda_E$ denotes the weight attributed to the bandwidth constraint whose value was assumed to be fixed and hence soft (Diamantaras et al., 1999). From Equation 2, it can be noted that the transfer function of the encoding filter’s spatiotemporal gain is inversely proportional to the magnitude of $\lambda_E$. Hence, larger values of $\lambda_E$ will lead to a proportionally smaller signal variance that is transmitted across the communication channel. As illustrated in Appendix B (Figures 7c and 7d), the transfer function of the encoder is generally low-pass when the assumed signal-to-noise ratio of transmitted signals is low but may be band-pass when high. From Equation 2, it can also be observed that the early gain control mechanism $G(\mu)$ will suppress the variance of input noise signals. This suppression necessarily allows signal-to-noise ratios to increase as a function of the visual signal’s mean thus justifying a luminance-dependent change in the transfer function of the encoding filter. We also allowed $\sigma_{tt}$, the standard deviation of the TAF, to linearly vary as a function of the root mean square (RMS) contrast of test signals. This assumption is easily justified because one would expect the RMS contrast of test signals to be proportional to $\sigma_{tt}/\mu$. The latter manipulation has two effects on the transfer function of the encoding filter: (i) the cut-off TF and (ii) the gain of the encoding filter will both increase (Appendix B). Both effects, when incorporated into the design of an adaptive encoder, lead to an adjustment (whitening) in its temporal transfer function that varies as a function of both the mean luminance and contrast of spatiotemporal signals (Mante et al., 2005).

(3) A decoding filter, $D(\omega)$, whose dual purpose is to invert the transformation imposed in the signal at the stage of signal encoding and reduce the contribution to the overall $MSE$ (de-noise) of the system that arises from the sources of uncertainty implied by the channel and input noise assumptions. The TF tuning of the decoder is also determined by the signal and noise statistics but unlike the encoder is generally low-pass thus reflecting the transfer
function of the TAF (Appendix B; Figures 7e and 7f). In line with previous researchers, the transfer function of the decoder was held constant (Atick, Li, & Redlich, 1993; Fairhall, Lewen, Bialek, & de Ruyter van Steveninck, 2001; Langley & Anderson, 2007). The step was taken as a minimal requirement that enabled the signal transmission model to explain the empirical data reported in this paper.

The combined transfer function of both the encoder and the decoder should, under global optimum settings, be low-pass (Equation A10; Figures 7g and 7h) so that those signals that are expected to have the largest contrast (low TF signals) are least attenuated by the signal transmission system. Visual perception, however, exhibits a band-pass TF contrast sensitivity function at high levels of mean luminance (Bowker, 1983; Camisa & Zemon, 1985; Georgeson, 1987). To explain the low TF loss by the luminance (Bowker, 1983; Camisa & Zemon, 1985; Georgeson, 1987). To explain the low TF loss by the luminance (Bowker, 1983; Camisa & Zemon, 1985; Georgeson, 1987).

General methods

Stimuli

Stimuli were generated in MatLab™ on a PC microcomputer using software adapted from the PsychToolbox routines (Brainard, 1997; Pelli, 1997) and were displayed with a GeForce4 MX440 graphics card at a frame rate of 75 Hz. Stimuli were presented either on a Sony 400PS 19-in. RGB monitor with a mean luminance of 40 cd/m² or on a Phillips BrightView 19-in. monochrome monitor with a mean luminance of 400 cd/m². The luminance gamma functions for each RGB color were separately measured with a Minolota CS100 photometer and were directly corrected in the graphics card’s advanced settings control panel to produce linear 8-bit resolution per color. The RGB monitor settings were adjusted so that the luminance of green was twice that of red, which in turn was twice that of blue. This shifted the white point of the monitor to 0.31, 0.28 (x, y) at 40 cd/m². A bit-stealing algorithm (Tyler, 1997) was used to obtain 10.8 bits (1785 levels) of luminance resolution on the RGB monitor under the constraint that no RGB value could differ from the others by more than one look up table step. RGB inputs to the monochrome monitor produced a gray scale image that was the linear sum of the RGB inputs and thus gave a 9.6-bit resolution (768 levels). The displays measured 36° horizontally (1152 pixels) and 27° vertically (864 pixels), and were 57 cm from the observer, in a dark room.

Stimuli were 4° fields of spatiotemporal white noise patterns presented for 853 ms (128 × 128 pixels × 64 frames). The TF spectrum of the noise was digitally filtered in the Fourier domain with log exponential filters:

\[ F(\omega) \propto \exp \left( -\frac{\ln(\omega/\Omega_p)^3 \ln 2}{(b_{0.5} \ln 2)^3} \right), \]  

where \( \Omega_p \) controlled the peak TF and \( b_{0.5} \) controlled the bandwidth, which was fixed at 1 octave. The SF spectrum of the white noise was unchanged. The noise was spatially windowed with a circular aperture whose edges were smoothed with a raised cosine envelope over 0.5°. The onset and offset of the noise were smoothed with a raised cosine over 40 ms (three video frames).

In a related work (Langley & Bex, in press), we have reported large contrast threshold elevations after only 1 s of adaptation to temporal contrast. To reduce adaptation effects between targets across trials, we presented the stimuli in one of four locations, at random across trials, each centered 2.8° from fixation (2° left or right and 2° above or below fixation). This resulted in an average of 3 s between any test stimuli, including the response time.
and intertrial interval. Furthermore, all TF conditions were randomly interleaved in a single run lowering the probability that similar TFs would appear in successive trials. Collectively, these experimental conditions minimized contrast adaptation effects in this study.

**Experiment 1**

TF discrimination

We first measured TF discrimination thresholds as a function of TF to estimate a series of single JND steps for each subject. The peak TF of a standard noise pattern was fixed at 1.2, 2.3, 4.7, 9.4, 18.5, or 37.5 Hz (1, 2, 4, 8, 16, or 32 cycles per 64 frame movie, respectively). The six standard TFs were randomly interleaved in a single run. The peak TF of a test noise image was the same as that of the standard, less a TF difference, $\Delta_T$, that was under the control of a staircase (Wetherill & Levitt, 1965). The staircase reduced $\Delta_T$ by 1 dB (1/20 log unit) following three correct responses and increased $\Delta_T$ by 1 dB after 1 incorrect response. The staircase was initialized with a $\Delta_T$ of 10% of the standard TF $T$ at random and terminated after 10 reversals or 50 trials, whichever occurred first. The observer's task was to fixate a central point then to indicate the location of the target with lower TF by pressing a corresponding button on a number pad. The fixation mark was a 3.75' square (2 × 2 pixels), whose color was used to provide feedback (green for correct, red for incorrect that was photometrically isoluminant with the background on the RGB monitor; and white for correct, black for incorrect on the monochrome monitor).

Although there is evidence for contrast constancy at low TFs, a reduction in STAC has been reported at high TFs (Georgeson, 1987). To eliminate this potential cue to the identity of the target pattern, we assigned the RMS contrast of the standard and the test patterns a value between 20% and 60% at random each trial. This forced observers to discriminate the noise stimuli based on their flicker rate rather than their physical or apparent contrast.

The raw data from a minimum of four runs for each condition (at least 180 trials per psychometric function) were combined and fit with a cumulative normal function by least $\chi^2$ 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). TF discrimination thresholds were estimated from the 75% correct point of the psychometric functions and 95% confidence intervals on these points were calculated with a bootstrap procedure, based on 1000 data sets simulated from the number of experimental trials at each test level (Foster & Bischof, 1991).

**Figure 2.** Temporal frequency (TF) discrimination functions for two observers (P.J.B. and K.L.) and three mean luminances, shown in the legend. The continuous curve shows a power function fit to the minimum threshold across subjects.

**Results and discussion**

Figure 2 shows TF discrimination thresholds for the authors. The data show that TF discrimination thresholds increase with TF in line with previous estimates given the presentation time of the test stimuli (McKee & Taylor, 1984). The continuous curve shows the best fitting power function to the most sensitive point (lowest threshold) across observers and conditions at each TF. This power function was used to generate a series of TF pairs separated by a maximum of 1 JND. Starting at 37.5 Hz (the highest TF that could be presented on our 75 Hz monitors), the power function was used to estimate the TF that was 1 JND lower. The next TF was 1 JND lower than that and so on until 1.2 Hz (one cycle per movie, the lowest nonzero TF in our 64-frame movie) was reached. By fitting the lowest threshold across observers and conditions, the power function represented the most conservative series that could be used for all observers and conditions. This facilitated comparison of data across subjects and conditions and produced 16 consecutive TF steps, a stationary pattern (0 Hz), and 15 pairs between 1.2 and 37.5 Hz. The filled circles on this function in Figure 2 identify the TFs of adjacent pairs for the match patterns that were actually used in Experiments 2 and 3.

**Experiment 2**

Threshold and suprathreshold contrast sensitivity

The purpose of the second experiment was to measure STAC as a function of the TF and contrast of the test signals. The TF discrimination functions of Experiment 1
were used to generate a series of flickering noise pairs that differed in TF by 1 JND or less. In plotting our results, we have transformed our data from the dimensionless units of contrast into apparent luminance modulations (LM) whose units are candela per square meter. This step was taken to ensure compatibility across the empirical data and model fits. Although this transformation is not essential for Experiment 1 where mean luminance was fixed, it is necessary when fitting the transmission model to the data for Experiment 3 because mean luminance is explicitly manipulated as a dependent variable.

Contrast/LM matches were obtained between all adjacent TF pairs to generate overall STAC as a function of TF. A standard and a reference pattern were presented in two of the four possible locations each trial. Four locations were used to increase the mean time between targets in the same locations and hence minimize adaptation effects. The RMS contrast of the standard noise was fixed at 0.1 (at 40 cd/m² mean luminance, LM depth = 4 cd/m²), 0.2 (LM = 8 cd/m²), 0.4 (LM = 16 cd/m²), or 0.6 (LM = 24 cd/m²). The RMS contrast of the match noise was under the control of a staircase (Wetherill & Levitt, 1965) that reduced the match contrast by 1 dB following two positive responses and increased it by 2 dB after one negative response. The staircase was initialized with a random contrast within ±3 dB of the standard contrast and was terminated after 10 reversals or 50 trials, whichever occurred first. The observer’s task was to fixate a 3.75° central square and to indicate the location of the target with higher apparent contrast by pressing a corresponding button on a number pad. The 15 match TF pairs were randomly interleaved in a single run and the raw data from a minimum of four runs for each condition (at least 180 trials per psychometric function) were combined and were fit with a cumulative normal function by least $\chi^2$ fit. Match contrast was estimated from the 50% point of the psychometric function, and 95% confidence intervals on this point were calculated with a bootstrap procedure (Foster & Bischof, 1991).

Contrast thresholds were measured in the same way as TF discrimination thresholds for flickering noise patterns in Experiment 1, except that the contrast of a single target was varied by the staircase. The observer’s task was to indicate which of four locations contained the flickering target with visual feedback at the fixation point. Target TFs were randomly interleaved in a single run and the data from a minimum of four runs were combined and fit as before. Contrast sensitivity was taken the inverse of the contrast threshold (75% correct), relative to the target with lowest threshold.

**Model fitting**

The transfer function of a signal transmission system (Figure 1) may be controlled by a number of parameters: the Lagrange multiplier $\lambda_E$ whose magnitude is determined by the channel constraint but was fixed in our simulations; the TAF, $R_{tt}(\omega)$, which is signal dependent and controls the tuning of the encoding filter; and the channel and the input noise variances ($\sigma_{ch}$ and $\sigma_{in}$, respectively), the latter of which we have assumed to be controlled by early retinal gain control processes. From our psychophysical experiments, we are unable to make definitive statements concerning the adaptability of each term employed by the signal transmission model. The main difficulty is because the net transfer function of a signal transmission system is principally controlled by the signal-to-noise ratio and the magnitude of channel constraints. Neither term can be constrained by our experiments. Given this difficulty, we allowed two of the model’s parameters to vary freely. One was the magnitude of $G(\mu)$, which we assumed affected the variance of input noise. The second was the variance ($\sigma_{tt}^2$) of the TAF for the encoding filter ($E(\omega)$) that, as already mentioned, one could expect to be proportional to the contrast of test signals. All other system parameters were estimated from each subject’s empirical data and then held fixed while the mean luminance and the contrast of our test signals were varied. For subject K.L., we collected 169 data points; whereas for subject P.J.B., we collected 142 data points. The signal transmission model (Equations 1 and 2) was fitted to each subject’s empirical data (reported in Experiments 2 and 3 simultaneously) in a single batch. In total, there were 7 parameters (noise variances, Lagrange multipliers) whose values were fixed throughout the simulations, 3 input noise parameters whose values were allowed to vary freely as a function of the mean luminance, and 9 further parameters to model the input signal variances for both the threshold contrast and suprathreshold contrast matching tasks (11 conditions).

Figure 3. (a and b) Threshold and suprathreshold apparent luminance modulation (LM)/contrast as a function of temporal frequency (TF) for two observers (P.J.B. and K.L.) with a mean luminance of 40 cd/m². The black data show LM sensitivity, which is expressed as the inverse of threshold LM relative to peak sensitivity. Colored data show iso-apparent LM functions for matches in which the absolute LM depth of the standard pattern is shown in the legend. Modulation depths of 4, 8, 16, and 24 cd/m² correspond to root mean square (RMS) contrasts of 10%, 20%, 40%, and 60%, respectively. The LM matching data are expressed as the LM depth relative to the flickering pattern with highest apparent LM. (c and d) Same as panels a and b but with the vertical axes plotted on a logarithmic scale. (e) Estimates of the signal temporal autocorrelation function (TAF) for the encoder ($\sigma_{tt}^2$, the single free parameter). (f) Transfer function of the system as a function of contrast with the fit parameters in panel c, sensitivity is normalized to thresholds for stationary (0 Hz) patterns. As LM/contrast increases, the transfer function of the transmission system whitens and the peak TF sensitivity increases.
Results and discussion

Figures 3a–3d show contrast sensitivity (black data) and relative STAC (colored data), expressed as LM depth, at RMS contrasts of 0.1 (LM = 4 cd/m², blue circles), 0.2 (LM = 8 cd/m², green squares), 0.4 (LM = 16 cd/m², red triangles), and 0.6 (LM = 24 cd/m², pink diamonds) for the two observers. Note also that Figures 3a–3b and 3c–3d show the same data but plotted on linear versus logarithmic ordinates. The STAC functions were generated by scaling adjacent TFs by the proportion of contrast gained or lost between each match pair. Any contrast change was then accumulated across the function. Note that one member of every match pair was fixed at the standard contrast/LM so the STAC functions show the overall change in apparent contrast with TF for a given reference RMS contrast. Although this process may accumulate a serial error with distance between any comparison points along the function, the sign of any such error is expected to be approximately equal on average. Furthermore, under contrast constancy, the function is expected to be flat. It was not possible to collect contrast matches at the highest TFs for the 60% contrast condition because the match tended toward impossibly high contrasts that produced look up table overflows.

Results shown in Figures 3a–3d are similar for both observers and show that contrast sensitivity is weakly band-pass, peaking at around 4 Hz, gradually falling at lower TFs, and sharply falling at higher TFs, in line with many previous studies (de Lange, 1958; Kelly, 1975; Roufs, 1972). STAC is also band-pass, even at the highest contrasts examined. Inspection of our data suggests that the magnitude of contrast constancy under the conditions examined here is smaller than is implied in previous research: the STAC matching functions do not become flat. The loss in STAC at TFs above around 8 Hz is in agreement with previous studies (Georgeson, 1987). It is useful to compare across Figures 3a–3b and Figures 3c–3d, where both subjects’ data have been plotted on a linear versus logarithmic vertical axes. The linear vertical axis (Figures 3a and 3b) more clearly shows a significant loss in STAC at lower temporal frequencies and high contrast signals of around 30% from the peak sensitivity. On a logarithmic vertical scale, that same loss in STAC (Figures 3c and 3d) at lower TFs is less apparent. Our data thus point to significant losses in STAC at low TFs that may have been less obvious in previous studies partly because of the logarithmic scaling of STAC matching functions.

We performed a variant of the contrast matching task to scale the relative heights of the functions across match contrast levels. In this paradigm, the RMS contrasts of the standard and the match images were fixed at different levels, and the observer adjusted the TF of the match image to match the STAC of the two patterns. The RMS of a standard flickering noise pattern was fixed at 10%, 20%, or 40% and was paired with a match pattern whose RMS contrast was fixed at 20%, 40%, or 60%, respectively.

The peak TF of the standard was fixed at 8 Hz, close to the peaks of the contrast sensitivity and STAC functions. The peak TF of the match increased under the control of a two-up two-down staircase, designed to converge on a TF that produced equal numbers of match contrast higher and match contrast lower responses. This process estimated a match TF at which a higher contrast pattern had the same apparent contrast as an 8-Hz standard of lower contrast. This process allowed us to estimate the relative heights of the functions, albeit crudely and suffering from the criticisms of this paradigm we have leveled above. Following this scaling, the STAC of the peak contrast (around 8 Hz) for each contrast matching function became approximately equal to the physical contrast of the noise pattern. This remarkable observation suggests that peak STAC increases approximately linearly with physical contrast, confirming earlier observations (Kulikowski, 1976).

The smooth curves drawn through Figures 3a–3d show the fits of the efficient transmission model provided an excellent fit to the data. From the model fits, the $R^2$ values were .991 for subject K.L. and .986 for subject P.J.B. Estimates for the rate parameter $\lambda$ were 3.2 and 2.9 Hz for subjects K.L. and P.J.B., respectively. Figure 3e shows the variance estimate of the TAF for the encoder ($\sigma_{\nu}^2$), which was the only free parameter in the model that was allowed to vary as a function of the contrast of the matching signals. Estimates for $\sigma_{\nu}^2$ insofar as model fitting was concerned were constrained to lie on a linear function whose slope and intercept was estimated from the data. In forcing this restriction on the variance parameter for the TAF, we confirmed that the peak in STAC functions linearly increases with physical contrast. Note also that the parameter $\sigma_{\nu}^2$ determines the shape of the encoding filter $E(\omega)$ and hence the net transfer function ($H(\omega)$) of the signal transmission system. The estimates for $\sigma_{\nu}^2$ linearly increased with the contrast of the test signals, which in an adaptive system can lead to a whitening of the transfer function of the STAC matching function. Estimates for the transfer function of the signal transmission system are shown in Figure 3f, which were required to explain the trends in subject K. L.’s data. From the figure, it can be seen that as the contrast of the test signal increases, the filter whitens—that is, the bandwidth increases and the filter becomes increasingly sensitive to high TFs. This observation confirms one of the predictions made by the transmission model; namely, that temporal bandwidth is controlled by underlying signal-to-noise ratios.

Experiment 3

LM, TF, and mean luminance

The spatiotemporal tuning of the signal transmission system is expected to vary with the signal-to-noise ratio,
being temporally low-pass (i.e., increasing temporal integration time to smooth noise) at low signal-to-noise ratios and becoming band-pass at high signal-to-noise ratios (Atick, 1992). This expectation is supported by the observation that at low light levels, temporal contrast sensitivity is relatively low-pass (de Lange, 1958; Kelly, 1961; Roufs, 1972; Snowden, Hess, & Waugh, 1995) and temporal integration durations shorten (Kahneman & Norman, 1961; Roufs, 1972; Snowden et al., 1995) and temporal sensitivity becomes band-pass as mean luminance increases, however, temporal contrast sensitivity becomes band-pass (de Lange, 1958; Kelly, 1961; Roufs, 1972; Snowden et al., 1995) and temporal integration durations shorten (Kahneman & Norman, 1964). In Experiment 3, we examine how a manipulation of the mean luminance affects contrast sensitivity and STAC functions. According to the proposed signal transmission model, we expect to observe an enhancement of whitening effects at the higher mean luminance levels owing to a possible reduction in input noise variance levels.

To test this idea, we repeated Experiment 2 using a flickering noise pattern with an RMS contrast of 10% or 40%. The mean luminance levels were set at 0.8 and 400 cd/m^2. Contrast thresholds were measured using the same procedures to those given in Experiment 2. The low mean luminance data were collected on the 40 cd/m^2 RGB monitor while the subjects wore neutral density filter spectacles (NoIR, Optima Low Vision services) with 2% transmission. To ensure stable levels of light adaptation, subjects were dark or light adapted for a minimum of 10 min before data collection and runs lasted in excess of 30 min. The high mean luminance data were collected on a Phillips Brightview monochrome monitor.

Figures 4a and 4c show contrast sensitivity (expressed as LM depth for ease of comparison across mean luminances), whereas Figures 4b and 4d show STAC as a function of TF and mean luminance (see legend) for two observers (P.J.B. upper and K.L. lower). The contrast sensitivity data for a mean luminance of 40 cd/m^2 (green squares, left) and STAC matches at RMS contrasts of 10% (green circles, right) and 40% (green squares, right) are replotted from Figures 3a to 3b for comparison purposes.

Results and discussion

At the RMS contrasts examined, both the contrast sensitivity and the STAC functions were low-pass at low mean luminance levels (blue data). As mean luminance increased, however, both contrast sensitivity and STAC functions became more band-pass as a function of TF. The contrast sensitivity results are in good agreement with previous data (de Lange, 1958; Kelly, 1961; Roufs, 1972), even allowing for differences in stimuli (flickering noise compared with homogenous fields). The STAC data again show mild evidence for contrast constancy. For each fixed mean luminance, peak TF sensitivity increased with RMS of contrast, as already noted in Experiment 2: At a contrast of 10%, the peak TF tuning shifted from around 2.5 Hz at 0.8 cd/m^2 to 8.0 Hz at 400 cd/m^2. At 40% contrast, the peak TF shifted from 5 Hz at 0.8 cd/m^2 to 10 Hz at 400 cd/m^2. The peak spatiotemporal sensitivity of the visual system thus increases as a function of both the mean luminance and contrast of the test signals.

Figure 5c shows how input noise estimates decreased as the test luminance increased, whereas Figure 5d shows the tuning of the encoding filter that was determined from the model fits to each subjects’ data shown in Figure 4. There is marked whitening—i.e., an increase in sensitivity to high TF signals as mean luminance increased from 0.8 to 40 cd/m^2. Both subjects subjectively noted a significant increase in the visibility of high TF stimuli at high mean luminances.

To quantify the effects of contrast constancy, we now refer to the model simulations, which are shown in Figures 3 and 4 by the continuous curves. In fitting the model, the ensemble of subjects’ STAC and the contrast sensitivity functions from both Experiments 2 and 3 were fit simultaneously using the signal transmission model outlined in the introductory sections. \( R^2 \) values from the model fits were, therefore, the same as those reported in Experiment 2. For variations in mean luminance, however, we allowed the standard deviation of the input noise \( \sigma_m \) to vary as a function of mean luminance to simulate the effects of early luminance gain control \( G(\mu) \) on an adaptive transmission system. As in Experiment 2, the variance of the autocorrelation function used to defined the encoding filter \( E(\omega) \) was allowed to vary as a function of the RMS contrast of the test signals.

On the whole, the curve fits achieved by the signal transmission model were effective in predicting the empirical data. From Figures 4a to 4c, it can be seen that the signal transmission model slightly overestimated STAC at high TFs at 0.8 cd/m^2 mean luminance (blue curves). The reason for the discrepancy is unclear. However, note that STAC could not be measured above 18 Hz for the 10% contrast condition. This was because the standard stimuli were at or near the contrast detection threshold. It is therefore possible that the practical difficulties when conducting matching tasks near to threshold underestimated STAC. As outlined by Brady and Field (1995) and Cannon (1979), near threshold stimuli may include a proportion of 0% contrast matches, lowering STAC overall. This may have affected the predictability of the signal transmission model for the low mean luminance and high TF conditions.

As in Experiment 2, the transfer function of the encoder could be modified by allowing the variance of the TAF to covary with RMS contrast. Figures 5a and 5b show for both observers the estimate of the variance of the TAF \( \sigma_m^2 \) for the encoding filter \( E(\omega) \), taken from the model fits for Experiments 2 and 3. The rate of increase in the variance of the encoding filter’s TAF itself increased with mean luminance.

Figure 5c shows the standard deviation of the input noise for the encoding filter, which was allowed to vary with the mean luminance of the image signal. The noise
variances of the input noise required to explain our data show a tendency for saturation at a relatively low mean luminance (somewhere between the 0.8 and 40 cd/m² used in this study). Figure 5d shows the resultant TF tuning of the sensitivity of the encoding filter. As luminance increases from 0.8 to 40 cd/m², the encoding filter became increasingly band-pass and its peak sensitivity shifted to the higher TFs with a relative loss in sensitivity at lower TFs. Interestingly, the peak TF of the estimated encoding filter’s transfer function was found to be around 18 Hz at 40 cd/m² mean luminance. This TF coincides with the peak in spatiotemporal adaptability of the visual system at threshold contrast levels for this mean luminance (Langley & Bex, in press).

In Figure 6, we have replotted results from Figure 5 as the ratio of both subjects’ (and model’s) contrast/LM matches taken across the 10% and 40% conditions for the three different luminance levels examined. In plotting the ratio of contrast matches, the figure reveals a more sensitive summary of the effects of filter whitening as a function of the contrast of test signals. At high mean luminances, where it was possible to detect high test TFs, the figures show that the relative spatiotemporal sensitivity for both subjects increased at the higher versus the lower TFs. The increases in sensitivity, especially at mean luminances of 400 cd/m² and temporal frequencies around 37 Hz, were found to be close to 50%. Note that if the above whitening effects are
translatable into effects on perception at threshold contrast levels (a fast adaptation mechanism), then one would expect that threshold elevations are temporal low-pass functions, which would be consistent with the predictions made by the adaptation of a sustained spatiotemporal channel. Sustained threshold contrast elevations according to a filter whitening hypothesis could lead to an adaptation-driven facilitation in threshold contrast: but only for the higher spatiotemporal frequency signals; a prediction that the authors’ have found some evidence to support (Langley & Bex, in press).

Figure 5. Model parameters for Figures 2 and 3 fits. (a and b) The variance of the encoding signal temporal autocorrelation function (TAF) as a function of luminance modulation (LM) (contrast) and mean luminance for the two observers. (c) The reduction in input noise variance required to explain changes in TF sensitivity as a function of the mean luminance. (d) The TF tuning of the encoding filter at threshold LM levels for the low and mid mean luminance conditions. The estimated transfer function of the encoding filter at contrast threshold levels did not change across the mid to high mean luminance conditions.

General discussion

Contrast constancy

Many previous studies have reported that STAC is relatively invariant of SF or TF (for a review, see Georgeson, 1990). If STAC were invariant of TF, as the constancy hypothesis dictates, then our matching functions should be approximately flat, at least over some proportion of the range of TFs examined in this paper. Our
data, however, suggest that both threshold and STAC strongly depend on TF at all mean luminances and contrasts studied. We do, however, find a modest increase in the relative STAC of high TF patterns in the direction of contrast constancy. This trend is predicted by a whitening of the TF bandwidth of the visual system that is both luminance and contrast dependent: a feature of our data that has not been addressed in previous empirical studies. The luminance dependency of constancy effects imply that the underlying mechanism responsible for constancy occurs quite early in the visual pathways (see also Heinrich & Bach, 2001; Mante et al., 2005).

We attribute the difference between our results and previous studies to two main sources. Firstly, close inspection of the data in previous studies that claim to have observed contrast constancy actually do show a clear dependence on SF or TF, and so the description contrast constancy is at best an approximation. For flickering patterns, our data are in line with previous studies that show a loss in STAC at high TFs (Georgeson, 1987). However, we also find that, like threshold contrast sensitivity (de Lange, 1958; Kelly, 1961; Roufs, 1972), there is a relative loss in STAC for low TF patterns. These findings support the objection that log-scaling the data gives only the appearance of STAC functions that look flatter than contrast sensitivity functions (Brady & Field, 1995; Cannon, 1979; Kulikowski, 1976).

The second and more important source of the difference is methodological. Observers report that it is difficult to match the STAC of images with very different appearances and they may therefore be influenced by other properties of the stimuli. Our technique, adapted from a previous study (Metha et al., 1998), avoids this difficulty and makes the task easy and less subjective to perform. All image pairs in our paradigm not only appear more similar than in previous studies, they cannot be reliably discriminated from one another on 25% or more trials. This removes some possibilities for subjects’ bias, from the a priori knowledge or expectation that the STAC of previously seen high TF patterns are low, which could introduce a compensatory response by subjects in a matching task.

In this study, the contrast of one of the match pair was fixed at a standard (10%, 20%, 40%, or 60% of the mean luminance), and the STAC function was accumulated from the sequence of relative contrast change across TF. The STAC functions are therefore based on matches between stimulus pairs in which one member was always presented at the target contrast. In the original study on which our paradigm was based (Metha et al., 1998), STAC functions were generated by starting at one end of the function and estimating STAC matches for each pair of spatial frequencies in turn, termed iso-apparent contrast contours. The STAC match functions for each SF became the standard for the next SF and so on until the other end of the SF range was reached. This procedure also showed a loss in STAC at high SFs and little evidence for contrast constancy. However, the loss of STAC at high SFs means that the physical contrast of match pairs tended toward low contrasts, where gain control (if present) may have generated intermediate levels of constancy. The modified paradigm used here ensured that one member of the match pair was always at the standard contrast, which avoids this possibility and also helped to minimize a risk that zero contrast matches may be included when the standard pattern was near its detection threshold (Brady & Field,
The paradigm employed here therefore reduced the possibility for an accumulation of errors along the STAC functions.

A further possibility for our failure to observe contrast constancy was suggested to us by Walter Makous (personal communication). The concern raised by Dr. Makous is that even if SF- or TF-dependent contrast gain control mechanisms were capable of supporting contrast constancy under typical viewing conditions, it might not be measured by our stepwise paradigm. Stepwise paradigms employ pairs of stimuli that are not discriminably different, indicating that the underlying channel responses are not reliably different (and thus do not support SF or TF discrimination). Under these conditions, gain control, which is based on relative channel activity, may not be different across channels and therefore contrast constancy effects may not be expected. However, our paradigm works by accumulating an overall trend from a large number of small effects, each of which may not be detectable in isolation, but nevertheless small gain differences must be expected to accrue. Furthermore, empirical evidence generally points to two (or possibly three) broadly tuned temporal channels with fixed bandwidth—for a review, see Watson (1986). With two broadly tuned temporal channels, it would not be possible to generate flat STAC functions by adjusting the gain of these filters without setting new local gains for all comparison pairs.

As mentioned in the Introduction section, Brady and Field (1995) have questioned whether gain control processes are necessary to explain the effect of contrast constancy. They reasoned instead that contrast constancy might be explained by intrinsic noise whose effects are dominant at low contrasts. In their experiments, Brady and Field (1995) examined STAC functions, but only at relatively low mean luminances (~40 cd/m²). Our experiments do not strongly disagree with their results, but we have examined effects of contrast constancy at higher and lower levels of mean luminance than they examined. At high mean luminances, our results show significant whitening (contrast constancy) but only at the highest range of spatiotemporal frequencies examined. Given the contrast range over which we have observed a trend toward constancy (between 10% and 60% contrast), it is difficult to see how a linear model could be capable of explaining our results.

Figure 3 shows that TF tuning depends on mean luminance and contrast: as mean luminance increase, the functions progressively become more band-pass and the location of peak sensitivity increases to higher TFs. This is consistent with previous estimates of temporal contrast sensitivity data (de Lange, 1958; Kelly, 1961; Roufs, 1972) and extends these earlier observations to supra-threshold contrast levels. In this study, we used band-pass-filtered noise patterns. The SF, orientation, and TF spectra of our stimuli are much broader than the sine grating patterns typically used in psychophysical research. It is possible that our stimulus selection has led to under-estimates of contrast constancy effects. The reason again draws upon the actual neural location at which constancy effects arise. If constancy effects can arise from later cortical processing, then one might predict greater constancy effects when testing with one-dimensional signals by virtue of the increase in signal power passed along orientation and SF tuned channels. One experiment to test this possibility might be to vary the orientation bandwidth of test signals to determine whether contrast constancy effects are orientation and SF tuned, thus pointing to a cortical locus. An orientation-tuned effect thus demonstrated would not require a significant change to our model. This is because a later stage of contrast constancy could be explained by an adjustment in the parameters of the TAF at the stage of signal decoding (extraction) as well as encoding. In this paper, we have applied Occam’s razor by explaining contrast constancy effects from an early stage in the transmission of spatiotemporal information as a minimal requirement that we found capable of explaining our data.

For static patterns, contrast constancy has been reported across changes in mean luminance but this requires a stable state of adaptation (Georgeson & Sullivan, 1975; Kulikowski, 1976; Peli, 1995), except at mean luminances below around 8 cd/m² (Peli, Yang, Goldstein, & Reeves, 1991; Peli, 1995). We did not compare contrast across mean luminance and our stimuli were binocularly viewed on homogenous fields; however, our data (with >10 min adaptation) and model are consistent with a tendency toward contrast constancy, with a relative loss in apparent contrast at high TFs, which favorably compares with the loss in apparent contrast at high SFs for static images (Peli, Arend, & Labianca, 1996).

**TF channels and tuning**

We have not explicitly specified the number or tuning of any underlying TF channels in our model. In principle, the overall transfer function of the system that we report could be supported by any number of TF channels, whose collective tuning and bandwidth changes are shown in Figures 3 and 5. On the basis of TF masking, discrimination and adaptation studies (e.g., Cass & Alais, 2006; Kulikowski & Tolhurst, 1973; Legge, 1978; Mandler, 1984; Mandler & Makous, 1984; Hammett & Smith, 1992; Hess & Plant, 1985; Hess & Snowden, 1992; Richards, 1979; Snowden et al., 1995), there are thought to be two TF channels, one low-pass with peak tuning around 1 Hz and one band-pass with peak tuning at 6–8 Hz. Data showing improved TF discrimination at high TFs and masking studies with low SF targets have been used as evidence for the existence of a third channel band-pass TF channel with peak sensitivity at around 11–15 Hz (Hess & Snowden, 1992; Mandler & Makous, 1984) or possibly up to 18 Hz (Kelly & Burbeck, 1987). Our data show...
whitening at higher TFs and contrasts, which suggests that contrast sensitivity to the high TFs has increased. We have explained this increase by (i) an early suppression of mean luminance whose by-product also reduces in input noise and (ii) a rapid increase in the variance ($\sigma^2_{tt}$) of the TAF. In both cases, the underlying rationale for the adaptive whitening of the visual system’s TF processes stems from the enhancement in the signal-to-noise ratios of visual signals.

The whitening process implies that discrimination thresholds could enhance at high TFs, without requiring additional TF channels (Hess & Plant, 1985). Equally, the whitening at high TFs reported here might also explain the phenomenon of motion sharpening (Bex, Edgar, Smith, 1995; Burr, 1980; Burr & Morgan, 1997).

That a reduction in input noise may be used to explain the high TF luminance-dependent whitening of the visual system is a strong assumption. Actually, the possible importance of de-noising visual signals at an early stage has been previously suggested by Laughlin (1994). The input noise assumption is not a critical element of our model. This is because the whitening could, in principal, occur through an increase in the TAFs variance parameter (see Equation 1) at virtually any stage along the neural pathways that propagate visual information. We are therefore open minded to the possibility for alternative accounts to the constancy effects that we report. The possibility for multiple explanations, given the experiments reported in this paper, raises one difficulty with psychophysical research; namely, that the psychophysicist is often limited to observations about the overall transfer function of the visual system. This limitation leads to an ambiguity in the interpretation of empirical data, especially when studying signal transmission models because of their cascaded (serial) transfer functions.

To illustrate the ambiguity, consider the effect of a hypothetical early retinal gain control mechanism that attenuated retinal signals as a function of their mean square statistics. The effects of such a mechanism could rescale both the mean and variance (contrast) of the early visual signals independently (Mante et al., 2005). This possibility is easy to demonstrate. Denoting $X(t)$ as a stationary temporal signal we write:

$$E[X^2(t)] = \sigma^2_{tt} + \mu^2,$$

where $E[.]$ is the expectation operator. An early gain control mechanism that maintained a constant mean square response would thus be observed to adapt to a signal’s mean ($\mu$) and variance ($\sigma^2_{tt}$). Thus, an early gain control mechanism whose attenuation is proportional to the mean square of the (early) retinal signals could, if rapid, also lead to an explanation for our empirical results. The rapidity might also be symptomatic of a pseudo nonlinear system (Langley & Anderson, 2007).

We stress again that one could adjust the transfer function of either or both the encoder and the decoder to explain the experimental data reported in this paper. The decision to fix the transfer function of the decoder was influenced by the recent research (i) of Mante et al. (2005), who found that cells in LGN of cat could be adapted by both the mean and the contrast of probe stimuli; and (ii) of Heinrich and Bach (2001), who studied effects of fast contrast adaptation in retina from VEP potentials. Heinrich and Bach concluded that enhancements in contrast processing are tuned for TF and fast, thus paralleling the constancy effects reported here. One could assume a known transfer function for the decoder from which an optimal encoder may be designed (see Equation A9). Given that an optimal signal transmission virtually employs the same signal information at both the stages of encoding and decoding, it is necessarily the case that fewer assumptions were made in our approach because we have been able to design a decoder capable of explaining our data rather than assume a convenient mathematical description.

Although left largely unspecified, the neural location of the decoder could be in visual area V1 of the cortex (Wilson, 1997). If so, and in also noting that spatial dependencies have been ignored in this paper for brevity, it would not be problematic to modify the proposed transmission model toward this end. Rather than transmit a faithful representation of propagated signals, which is the goal of the transmission system developed in this paper, one could supplement that faithful representation with additional processes (filtering operations) whose ultimate design satisfies a different purpose (e.g., the detection of visual motion and/or spatial orientation). The key steps proposed here to explain constancy effects need not require modification. This is because the optimization of a signal transmission system assumes that one possesses knowledge of about the systems constraints, the underlying signal, the sources of signal uncertainty, and the purpose (target) of the system (Langley & Anderson, 2007). With these considerations, adjustments of an adaptable and an optimized system are necessarily determined by the underlying signal-to-noise ratios but subject to the ensemble of constraints imposed upon the system by the available resources at hand.

The spatiotemporal CSF of the visual system changes from low-pass to band-pass as a function of the mean luminance of the image (Bowker, 1983; Kelly, 1975; Koenderink & van Doorn, 1979; Robson, 1966; Snowden et al., 1995). The band-pass tuning of the CSF implies a relative loss in the lower spatiotemporal frequencies at the higher mean luminance levels (Snowden et al., 1995). Atick (1992) and Dong and Atick (1995) explained the band-pass tuning of the CSF with the idea that the visual system de-correlates spatiotemporal signals in combination with quantum (signal-dependent) noise assumptions. Our difficulty with a purely de-correlating encoding strategy arises from two arguments. First, an optimal
encoder should at the very least be informed about the transfer function of the decoder (Equation A9). Second, in the absence of intrinsic channel noise, the optimal encoding filter \( E(\omega) \) need only smooth temporal signals to reduce input noise for low signal-to-(input) noise ratios. For high signal-to-noise ratios, the same system need only scale (multiplicatively suppress) signals in order that they may be squeezed through a communication channel and recovered. According to the single channel encoding model illustrated in Figure 1, such an encoding scheme predicts that the net temporal transfer function of the visual system would be low-pass, which contradicts the band-pass temporal CSF that is observed empirically. A transient encoding scheme, however, can be justified under the circumstance where there is a constraint on the propagation of signals along the communication channel, in combination with channel noise. In the event of a noisy communication channel, the objective of the encoding filter should maximize the signal-to-noise ratio of the signals transmitted along the communication channel: a leaky-predictive coding strategy. By leaky, we refer to the principal of partial de-correlation (Langley, 2004; Webster, 1996). A global optimal leaky predictive-encoding strategy cannot, however, explain the transient threshold contrast functions reported here and by many other researchers (de Lange, 1958; Kelly, 1961; Roufs, 1972; Snowden et al., 1995). This is because the white assumptions for sources of signal uncertainty, in combination with the low-pass \( (1/\omega) \) amplitude spectrum attributed to natural scenes (Dong & Atick, 1995), imply that a decoding filter could be introduced into the visual pathways that will invert any encoding transformation thus leading to a low-pass CSF. To explain the loss in visual sensitivity at low TFs, we have assumed, like others (Atick et al., 1993), that there is a loss in information transmitted across a (neural) communication channel. The loss in this paper, arising from the assumption of a fixed decoding filter and source of signal uncertainty that we have posited to exist between the encoding and the decoding transformations.

We also fixed the magnitude of Lagrange multiplier that controls the channel constraint \( \lambda_E \) in Equation 2 and estimated its value from the empirical data. There were three reasons why this step was taken: (i) contrast discrimination thresholds do not linearly increase with contrast, but rather flatten at high test contrasts (Ross, Speed, & Morgan, 1993), which implies that the communication channel constraint cannot be viewed as hard (see Diamantaras et al., 1999); and (ii) one’s perception of contrast is approximately a linear function of increasing contrast in unadapted conditions (Kulikowski, 1976) and Experiment 2; and (iii) the “channel constraint” that best represents information transmission by the visual system is unknown.

In a companion paper (Langley & Bex, in press), we have shown that the spatiotemporal adaptability of the visual system peaks around 18 Hz. Temporal masking studies, however, report that the peak sensitivity of the visual system lies in the range of 10–12.0 Hz (Cass & Alais, 2006; Hess & Snowden, 1992). To explain the differences between masking and adaptation studies, we have proposed a similar model of spatiotemporal processing to the one illustrated in Figure 1. By placing a sustained decoding process in cascade with a transient encoder, it is necessarily the case that the net spatiotemporal transfer function of the visual system will be shifted to lower TFs, which could explain the difference between the peak in spatiotemporal adaptability and the threshold sensitivity function of the visual system. It is therefore of interest to note that the estimates for the encoding filter (Figure 5d) exhibit a similar peak at around 18 Hz (see also Mante et al., 2005). Finally, we stress that the single TF model illustrated in Figure 1 is not intended to challenge existing two-temporal models of spatiotemporal vision. Our choice of a single channel system was one that was chosen as a minimal requirement that we have found capable of explaining the empirical data that we have collected while at the same time provide insight into the underlying strategy that we believe are responsible for the tendency toward contrast constancy at higher levels of luminance and contrast.

Given that most previous research have employed visual display devices whose mean luminance lie in the range of 20–80 cd/m², with the exception of Snowden et al. (1995) who used a 140 cd/m² monitor, our results imply that the full range of TF and possible SF tuning of the visual system has been underestimated in laboratory conditions. In natural scenes, where mean luminance signals routinely exceed 1000 cd/m² (Martin, 1983), our results raise the possibility of considerable contrast sensitivity and adaptability in the 50- to 60-Hz range: Note that this corresponds to the frequency of flicker of many artificial illuminating devices. If so, controlled investigation of the spatiotemporal contrast sensitivity of the visual system under natural lighting conditions may prove useful to researchers of normal and abnormal visual function in clinical settings (e.g., Skottun, 2001).

### Summary

A stepwise paradigm was employed to measure STAC as a function of the TF, the absolute contrast, and the mean luminance of dynamic noise images. Our results show that STAC is highly TF dependent, with a loss in sensitivity at low and high TFs that depend on contrast and mean luminance. There is a tendency toward TF invariance (contrast constancy) at high mean luminances and contrasts where sensitivity to high TFs increases. These observations are explained by a signal transmission model of spatiotemporal processing in which a tendency toward contrast constancy emerges from spatiotemporal
whitening of an encoding TF filter. The whitening effects as explained in this paper shift peak sensitivity to the higher TFs, which increases sensitivity to high TFs through the accommodation of increases in underlying signal-to-noise ratios. Given the short presentation times of our test stimuli, the whitening effects imply that these processes are rapid.

**Appendix A**

In the appendices, we derive key equations required for the main part of the paper.

Let \( Y(\omega) = D(\omega)[E(\omega)X(\omega) + \varepsilon_{in}] + \varepsilon_{ch} \) refer to observations of the transmitted signal \( X(\omega) \), and let \( \varepsilon_{in} \) and \( \varepsilon_{ch} \) refer to the input and the channel sources of signal uncertainty, respectively. The transfer function of both the encoder \( E(\omega) \) and the decoder \( D(\omega) \) that minimizes the MSE between the input and observed signals is to be determined. Given the cascaded placement of \( D(\omega) \) \( E(\omega) \), the signal transmission problem is ill-posed, unless additional constraints are introduced. The additional constraint we assume to be penalty incurred by the transmission of high variance signals whose weight is given by the Lagrangian \( \lambda_E \). In placing this constraint on the variance of transmitted signals, the computational objective is to design optimal encoding and decoding filters that will deliver a minimum MSE in the respect of the input and output signals subject to a system constraint that imposes a restriction on the variance of signal propagated along the communication channel.

We let \( R_n(\omega) \) denote the TAF, with \( P_a \) referring to the maximum permissible variance that may be transmitted across the communication channel. Note that if \( \lambda_E \) is fixed, \( P_a \) can be ignored. Collectively, the MSE for the transmission system plus constraints may be described by the functional F (Franks, 1969):

\[
F = \int [X(\omega) - Y(\omega)]^2 + \lambda_E \left[ E^2(\omega)(R_n(\omega) + \sigma_{in}^2) - P_a \right] d\omega.  \tag{A1}
\]

To solve the constrained optimization problem given above, we differentiate the functional \( F \) under the integral with respect to \( D(\omega) E(\omega) \), which gives two stationary conditions:

\[
\frac{dF}{dD(\omega)} = 2D(\omega)E^2(\omega) \left( R_n(\omega) + \sigma_{in}^2 \right) - 2E(\omega)R_n(\omega) + 2D(\omega)\sigma_{ch}^2,  \tag{A2}
\]

\[
\frac{dF}{dE(\omega)} = 2D^2(\omega)E(\omega) \left( R_n(\omega) + \sigma_{in}^2 \right) - 2D(\omega)R_n(\omega) + 2\lambda_E E(\omega) \left( R_n(\omega) + \sigma_{in}^2 \right),  \tag{A3}
\]

in setting both derivatives to zero, we multiply Equation A2 by \( D(\omega) \) and Equation A3 by \( E(\omega) \). Subtracting the resulting equations gives

\[
D^2(\omega)\sigma_{ch}^2 = \lambda_E E^2(\omega) \left( R_n(\omega) + \sigma_{in}^2 \right).  \tag{A4}
\]

Equation A4 may be plugged into Equations A2 and A3 giving

\[
[E^2(\omega)]^+ = \left( \frac{R_n(\omega)}{R_n(\omega) + \sigma_{in}^2} \right) \sqrt{\frac{\sigma_{ch}^2}{\lambda_E R_n(\omega) + \sigma_{in}^2}} - \frac{\sigma_{ch}^2}{R_n(\omega) + \sigma_{in}^2},  \tag{A5}
\]

\[
[D^2(\omega)]^+ = \left( \frac{R_n(\omega)}{R_n(\omega) + \sigma_{in}^2} \right) \sqrt{\frac{\lambda_E R_n(\omega) + \sigma_{in}^2}{\sigma_{ch}^2}} - \lambda_E.  \tag{A6}
\]

From Equation A5, assume that the TAF \( R_n(\omega) \gg \sigma_d^2, \sigma_{ch}^2 \). This gives

\[
[E^2(\omega)]^+ \sim \sqrt{\frac{1}{R_n(\omega)}},  \tag{A7}
\]

and shows for high signal-to-noise ratios that the optimal encoder reciprocates the TAF.

Starting from Equation A1, assume that the encoder \( E(\omega) \) is known in advance but subject to an additional level of uncertainty denoted by \( \delta E(\omega) \). The observed signal is now given by \( Y(\omega) = D(\omega) \left[ (E(\omega) + \delta E(\omega))X(\omega) + \varepsilon_{in} \right] + \varepsilon_{ch} \). The optimal decoder \( D(\omega) \) in which the encoder is assumed known is found by the minimization of the functional \( G \):

\[
G = \int [X(\omega) - Y(\omega)]^2 d\omega,
\]

which gives the Wiener filter:

\[
D(\omega) = \frac{E^*(\omega)R_n(\omega)}{E^2(\omega) \left[ R_n(\omega) + \sigma_{in}^2 \right] + \sigma_{ch}^2 + [\delta E^2(\omega)R_n(\omega)]},  \tag{A8}
\]
Starting again from Equation A1, assume that the decoder $D(\omega)$ is known and fixed but that the encoder $E(\omega)$ can adjust. If the encoder cannot propagate changes in its transfer function to the decoder, but is still subject to a transmission constraint, the optimal encoder is found by differentiating Equation A1 with respect to the encoder $E(\omega)$ only. This gives

$$E(\omega) = \frac{D^*(\omega)R_u(\omega)}{D^2(\omega)[R_u(\omega) + \sigma_m^2] + \lambda_E R_u(\omega)} ,$$

and again assuming that the TAF $R_u(\omega) \gg \sigma_m^2$, $\sigma_{ch}^2$ one obtains

$$E(\omega) \sim \frac{D^*(\omega)}{D^2(\omega) + \lambda_E} ,$$

(A9)

so the optimal encoder approximately reciprocates the transfer function of the decoder under the circumstance where the signal-to-noise ratio is high.

Finally, the transfer function of the signal transmission system defined by the product $H(\omega) = \tilde{E}(\omega) D(\omega)$ from Equations A5 and A6 is given by

$$H(\omega) = \max \left\{ \sqrt{\frac{R_u(\omega)}{R_u(\omega) + \sigma_m^2}}^2 - \frac{\sigma_{ch}^2 \lambda_E}{R_u(\omega) + \sigma_m^2} \right\} ,$$

(A10)

which is a low-pass function.

Appendix B

Here we illustrate properties of the encoding and decoding filters given in Equations A5 and A6. In the illustrations, the noise variances, the channel constraint, and the signal’s mean were held constant.

The temporal encoder from Equation A5 may be thought to be composed of three terms. The far right term, $\frac{\sigma_{ch}^2}{R_u(\omega) + \sigma_m^2}$, represents the channel noise to encoded signal plus input noise variance ratio. Its main effect is to control the cut-off TF of the encoding filter. The far left term, $\frac{R_u(\omega)}{R_u(\omega) + \sigma_m^2}$, represents the ratio of the underlying signal to encoded signal plus input noise variance. Its effect is to attenuate the gain of the encoder under circumstances where the signal-to-noise ratio is low. The reason is because a lower transmission $\text{MSE}$ may be obtained under low signal-to-input noise ratios by first smoothing the encoded signal prior to transmission. At higher signal-to-input noise ratios, the transfer function of the encoding filter may broaden in spatiotemporal transfer function or whiten. These ideas lead to the following simple principle:

i. An optimal signal transmission system should if adaptable adjust its transfer function such that the degree of smoothing depends upon the underlying signal-to-noise ratio: As the signal-to-noise ratio increases, so may the level of smoothing should decrease (Atick, 1992; Langley, 2005; van Hateren, 1992).

The final expression that defines the encoder is given by $\sqrt{\frac{\sigma_{ch}^2}{\lambda_E (R_u(\omega) + \sigma_m^2)}}$. This function controls the shape of the encoder under the circumstance where the signal-to-noise ratio is high. Whence

ii. The transfer function of an optimal spatiotemporal encoder should (subject to principle i) approximately reciprocate the TAF (Dong & Atick, 1995).

Figure 7c illustrates the dependency of the encoder’s transfer function on $\sigma_u^2$, the variance of the input signal. When the signal-to-noise ratio is low, the encoder’s gain is low and its transfer function low-pass. As the signal-to-noise ratio increases, two changes to the transfer function of the encoder take place: (a) its gain increases and (b) its transfer function becomes more band-pass (see principles i and ii). Figure 7d shows the dependency of the encoder’s transfer function on the rate parameter $A$. Because the magnitude of $A$ is determined by the bandwidth of the underlying signal, larger values for $A$ imply a signal whose bandwidth is broader and hence the transfer function of the optimal encoder should also broaden to take into account this additional information.

The third component of the signal transmission problem is the recovery filter which we denote by $D(\omega)$. The decoder should

iii. invert the transformation made at the stage of signal encoding (subject to principles i and ii) and recover signals subject to the signal uncertainty from both the input and channel sources of noise, respectively.

The decoder obtained by jointly optimizing the transmission system is given by Equation A6. Figures 7e and 7f show the optimal decoder for the signal encoders given in Figures 7c and 7d, whereas Figures 7g and 7h show the net transfer function $H(\omega)$ of the signal transmission system (Equation A10). From the figures, two points can be noted: (i) as the variance of the underlying signal increases, the gain of the system increases in combination with a broadening (whitening) of the system’s spatiotemporal response; and (ii) as the rate parameter $A$ increases, a broadening in temporal bandwidth is also observed.
From Figures 7g and 7h, it should also be noted that the net transfer function \( H(\omega) \) of the signal transmission system is low-pass.

The optimal transfer function of both the encoding and the decoding weights depends upon the noise statistics, the signal statistics, the channel constraint, and the properties of the transmission pathways. Given the joint dependency of the system’s elements, one concludes that

iv. A global optimal transmission system should jointly adapt the transfer functions of both the encoder and the decoder as the statistics of the propagated signals change but always subject to the satisfaction of the channel constraint.

### Appendix C

#### Contrast thresholds

The signal discrimination function for the model considered can be derived from signal detection theory (Kontsevich, Chen, & Tyler, 2002), which gives

\[
d'(t) = \frac{\hat{x}_1(t) - \hat{x}_2(t)}{\sqrt{\text{var}[\hat{x}_1(t)] + \text{var}[\hat{x}_2(t)]^{0.5}}}
\]  

(C1)

as the difference in signal that is required to discriminate observations of a signal to a level determined by \( d' \) (assumed fixed). In Equation 8, \( \hat{x}_1(t) \) and \( \hat{x}_2(t) \) refer to two independent samples of the transmission system’s output with \( \text{var}[] \) denoting a variance operator. At threshold signal levels, we may set \( \hat{x}_2(t) = 0 \) and note

\[
E[h(t) * x_1(t)] = E[\hat{x}_1(t)]
\]

(C2)

where \( E[.] \) and * denote the expectation and convolution operators, respectively; \( h(t) = e(t) * d(t) \) is used to refer to the transfer function of the signal transmission system in the signal domain. The definition for the signal threshold \( t \) follows from the Fourier transform of Equation 9 in which we represent the Fourier transform of a probe signal by an impulse function at its peak tuning frequency \( (\omega_p) \), which gives

\[
T(\omega_p)A^2 \frac{\sigma_{\text{tot}}}{H(\omega_p)} = \sigma_{\text{obs}}^2 + \sigma_{\text{in}}^2 = \int H^2(\omega) d\omega
\]

\[
+ \sigma_{\text{ch}}^2 \int D^2(\omega) d\omega
\]

(C3)

as the one employed in this paper. The definition shows that signal thresholds are proportional to the standard deviation of the internal noise passed through the signal transmission system but inversely proportional to its sensitivity. If the sensitivity of the visual system’s transfer function reduces at the higher temporal frequencies, one would expect signal thresholds to increase (Brady & Field, 1995).

### Signal matching function

In letting \( \hat{x}_1(t) \) denote the estimate of an observed signal in one condition and \( \hat{x}_2(t) \) denote the estimate of the same signal in a second condition, a perceptual equality is given by

\[
\hat{x}_1(t) = \hat{x}_2(t),
\]

where the estimates for the two signals we have modeled by the convolution of the encoding and decoding operators which gives

\[
e_1(t) * d_1(t) * \hat{x}_1(t) = e_2(t) * d_2(t) * \hat{x}_2(t)
\]

as the model of the contrast matching function obtained from psychophysical experimentation. Assume that one can interchange contrast (at a fixed level of mean luminance) with the standard deviation of the observed signal, and that the observed signal is sufficiently concentrated about its peak in signal power so that it can be approximated by an impulse function at the peak. With this assumption, it is simple to take the Fourier transform of the above to give

\[
\frac{H_1(\omega_p, \sigma_{\text{tot}}, \mu)X_1(\sigma_{\text{tot}})}{X_2(\sigma_{\text{tot}})} \approx H_2(\omega_p, \sigma_{\text{tot}}, \mu),
\]

(C4)

where \( \omega_p \) represents the TF of the reference signal. The equation shows that the ratio of the contrast matches approximates the relative changes in the transfer function of the visual system as a function of the TF of the test signals, but only to the extent that the TF transfer function of the visual system does not significantly vary as a function of the variance of the underlying signal. In comparing Equation C3 with Equation C4, it can be seen that both signal thresholds and signal matching function depend upon the underlying sensitivity of the visual.

Figure 7. Shows the temporal autocorrelation function (TAF) with (a) \( A = 2 \) as a function of temporal frequency (TF). The different curves show that the TAF is scaled by the variance \( \sigma_{\text{tot}}^2 \) and (b) \( \sigma_{\text{tot}} = 10 \) for different values of the rate parameter \( A \). Notice that \( A \) controls the bandwidth of the TAF. (c and d) The transfer function of an optimal spatiotemporal encoder for different signal variances and the rate parameter \( A \). The soft-constraint \( \lambda_2 \) was assumed to be soft and thus held constant. (e and f) The corresponding decoders obtained from panels c and d and the net transfer function of the transmission system are shown, respectively. (g and h) In all cases, the figures show that increasing either the variance or the rate parameter of the TAF may broaden the spatiotemporal bandwidth of an optimized signal transmission system.
system’s temporal processors of visual information. In the case of contrast thresholds, the variance of the intrinsic noise sources is largely unknown and must therefore be estimated from the empirical data.

**Appendix D**

One derivation for the TAF employed in this paper follows from the random step function (Franks, 1969), which has been used to model the TAF for television signals (see Figure 8). Given that the number of change points depends only on the length of an interval, stationary conditions can be assumed for which

\[
P(0) = e^{-A|r|},
\]

and

\[
P(k) = \frac{1}{k} (A|r|)_k e^{-A|r|},
\]

with the TAF given by

\[
R_{tt}(\tau) = \sum \left[ x(t) x(t+\tau) \right] - \sum \left[ x(t) x(t+\tau) \right] + \sum \left[ x(t) x(t+\tau) \right]
\]

\[
= P(\text{one change point in } \tau) \bar{a}^2 + P(\text{no change point in } \tau) \bar{a}^2 - (1 - e^{-A|r|}) \bar{a}^2 + e^{-A|r|} \bar{a}^2
\]

\[
= \sigma_a^2 e^{-A|r|} + \bar{a}^2,
\]

whose Fourier transform, with \( \mu^2 = \bar{a}^2 \) and \( \sigma_a^2 = \sigma_{tt}^2 \), gives the desired result

\[
R_{tt}(\omega) = \mu^2 \delta(\omega) + \frac{\sigma_{tt}^2}{A \left(1 + \frac{\omega^2}{A^2}\right)}.
\]

**Acknowledgments**

We thank Dr. Walter Makous and two anonymous reviewers for helpful comments on this work. This research was supported by the Wellcome Trust and EPSRC Grant # EP/E030793/1.

Commercial relationships: none.

Corresponding author: Peter Bex.

Email: peter.bex@schepens.harvard.edu.

Address: Schepens Eye Research Institute, Harvard Medical School, 20 Staniford Street, Boston, MA 02114, USA.

**References**


