ELSEVIER

Physica D 148 (2001) 136–146

PHYSICA D

www.elsevier.com/locate/physd

Fractal time and 1/*f* spectra in dynamic images and human vision

Vincent A. Billock*, Gonzalo C. de Guzman, J.A. Scott Kelso

Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, FL 33431, USA

Received 17 January 2000; received in revised form 14 April 2000; accepted 21 August 2000
Communicated by R.P. Behringer

Abstract

Many physical and biological systems have $1/f^{\beta}$ Fourier spectra — a fractal attribute implying multiple similar mechanisms operating at various spatial and temporal scales. These scaling laws of physical phenomena should have correlates in perceptual mechanisms that have evolved to transduce them. We show that measures of a changing visual environment and perceptual measures of how we see it exhibit fractal-like multiscale characteristics; both dynamic images of natural scenes and human temporal frequency perception display commensurate $1/f^{\beta}$ spectral behavior. © 2001 Published by Elsevier Science B.V.

PACS: 87.10.+e; 42.66.Si; 47.53.+n

Keywords: 1/*f* Natural image statistics; Correlational structure; Temporal frequency channels; Human temporal vision

1. Introduction

At first blush, the visual world is diverse and complex. An image on the retina may move at many velocities over a short period of time and objects comprising an image may move independently, or be occluded by movement. In addition to object motion, observer motions (head, eye and locomotion) take place on many time scales, as do modulations of the illuminant. This kind of spatially and temporally multiscale physical system is often best described in the Fourier domain by $1/f^{\beta}$ spectra [1–11]. Likewise, 1/*f* characteristics, from the kinetics of ion channel membranes to the dynamic behavior of neural ensembles are ubiquitous in biology [12–17]. Since organisms must adapt flexibly to an environment in continuous flux, here we ask whether measures of a changing world (as described by a sequence of visual images) and perceptual measures are also 1/*f* in nature. To find out, we analyzed the dynamic visual environment by digitizing and Fourier transforming videotape scenes ranging from children chasing bubbles to windswept rainy rooftops. We compared these results to scaling relationships inferred from fractal correlation analysis of our data on human sensitivity to luminance contrast at many specific spatial and temporal scales. We show that dynamic images of natural scenes and human temporal contrast sensitivity exhibit commensurate $1/f^{\beta}$ behavior.

* Corresponding author. Present address: Logicon Technical Services, Inc., US Air Force Research Laboratory, PO Box 317258, Dayton, OH 45437-7258, USA. Tel.: +1-937-255-8879; fax: +1-937-255-8709.
E-mail address: vince.billock@wpafb.af.mil (V.A. Billock).

0167-2789/01/$ – see front matter © 2001 Published by Elsevier Science B.V.
PII: S0167-2789(00)00174-3

2. Temporal frequency spectra of dynamic natural scenes

2.1. Background: the spectral characteristics of natural images

Despite their apparent diversity, natural images [1] have surprisingly regular spectra; most images have Fourier spatial frequency spectra whose amplitudes follow the power law $A(f_s) = k/f_s^b$, where f_s is spatial frequency (generally measured in cycles/picture width or cycles/degree of visual angle). These findings are surprisingly robust; studies using different image selection criteria, and different photographic, digitization, and analysis methods all obtain strikingly similar results [5–11]. The value of b is generally between 0.7 and 1.4 [5–11]; a recent survey of the literature found that for 1176 images with published statistics, the mean value was 1.08 [18]. One factor in these spectra is the $1/f$ spectra of each edge in an image, another is the fractal-like multiscaled self-similarity of many natural images [19,20]. The situation for temporal spectra is not as clear. Since temporal frequency (cycles/s) = velocity (degrees/s) × spatial frequency (cycles/degree), one would expect that an image with $1/f_s$ spatial frequency spectra moving at constant velocity will give rise to a $1/f_t$ temporal frequency spectrum. However, images consist of many objects, each with its own image statistics and these objects may move independently, temporarily occluding one another, or disappearing and reappearing across the boundaries of the image. Illumination changes and eye movements (both pursuit and search) add complications as well. Compared to static images, there is surprisingly little data on this point. Some studies [21,22] have fit dynamic image spectra to lowpass functions that are $1/f$ at high temporal frequencies (see Section 2.3). Two other recent studies have examined some dynamic scenes and fit $1/f^a$ models to them [23,24]. Given the relative paucity of data on this point, we decided to examine the spectra of several dynamic video sequences.

[1] Definition of a natural image varies between studies but generally include landscapes, woodland scenes, foliage, etc. and exclude man made objects.

2.2. Characterization of dynamic image spectra: methods

Five videotaped sequences were chosen for analysis based on their variety of light source modulations and image movements (Fig. 1) including camera movements, weather (moving leaves, windswept rain), and especially biological movements by one or more humans (see Table 1 and Fig. 1 for brief

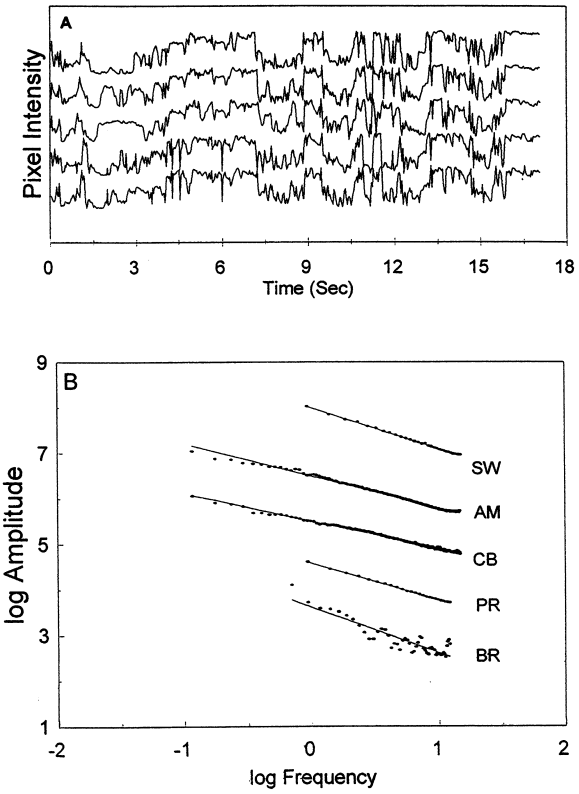

Fig. 1. Data on dynamic natural image sequences. (A) Examples (vertically offset) of pixel intensity time series followed over 512 frames (17 s) for five nearby pixels in the film sequence AM. (B) $1/f^{\gamma}$ amplitude spectra (vertically offset), for several motion picture scenes chosen to represent a variety of object and camera movement, as well as some light source modulation. The scenes are: PR, soldiers moving stealthily through a jungle (γ of 0.81). Sequence PR was taken from video of the film *Predator* (Twentieth Century Fox, 1987, J. McTiernan, dir.). BR, a policeman searching a windswept rainy rooftop in flickering light (γ of 1.02). Sequence BR was taken from video of the film *Blade Runner* (Warner Bros., 1982, R. Scott, dir.). SW, students walking to class (γ of 0.87). CB, children chasing erratically drifting bubbles (γ of 0.61). AM, a child climbing on playground equipment (γ of 0.68).

Table 1
Statistics of motion pictures[a]

Description	Pixels $H \times V$	Range (Hz)	$\gamma \pm$ S.D.	r^2
PR: soldiers stealthily moving through jungle	720×243	0.94–12	0.811 ± 0.005	0.999
BR: policeman searching rainy rooftop in flickering light	720×243	0.75–12	1.018 ± 0.050	0.904
SW: students walking to class	720×243	0.94–15	0.865 ± 0.013	0.990
CB: children chasing bubbles	320×240	0.115–15	0.608 ± 0.007	0.985
AM: child climbing on playground equipment	320×240	0.115–15	0.683 ± 0.006	0.991

[a] γ is the exponent of $f^{-\gamma}$ function fit to the amplitude spectra of an image sequence.

descriptions). The video sequences CB and AM were digitized frame-by-frame on a Macintosh AV1800 with built-in frame grabber. The digitizer's resolution was $320 \times 220 \times 8$ bits grayscale and 512 frames (representing image sequences sampled at 30 frames/s) were acquired (these sequences were digitized frame-by-frame). The video sequences SW, BR and PR were digitized on-the-fly with an Iris Indigo system; the digitizer's resolution was $720 \times 223 \times 8$ and number of frames was dependent on the computer's available memory. Spatial resolution per se is not important to our (primarily) temporal analysis, but each pixel provides an entire time series for analysis. The Fourier transform of a single 1D series is an unreliable predictor of the spectra, but the reliability of thousands of averaged spectra is good [25–27]. Each pixel's gray-level time series (Fig. 1A) is analyzed by fast Fourier transform (FFT) and the resultant spectra (160,560 individual pixel spectra each of for PR, BR and SW; 70,400 for AM and CB) are averaged (Fig. 1B). The sequences taken from videotapes made from commercial films (i.e., sequences BR and PR) have a Nyquist frequency of 12 Hz, compared to 15 Hz for directly recorded videotape (i.e., sequences SW, AM, CB). The minimum temporal frequency is determined by the sequence length (two periods of data were required).

2.3. Results and analysis

All five video sequences had temporal amplitude spectra that were well fit (except for sequence BR, $r^2 > 0.98$ for all fitted spectra) by $f^{-\gamma}$ functions ($0.61 \leq \gamma \leq 1.02$; see Table 1). Previous studies suggested that the correlation between frames (the autocorrelation function for the sequence) may drop exponentially as a function of frame separation [28,29]. The Fourier transform of this correlation function yields a power spectrum of $P(f) \propto k/(k^2 + f^2)$ (which approximates a $1/f$ amplitude spectrum for $f \gg k$). Other studies fit this spectrum derived from the exponential model to short movie scenes [21,22] but based on their published figures and an examination of their tabulated "k" parameters, a simple $f^{-\gamma}$ model would also fit most of their spectra. Our data also shows hints of a shallower spectrum at the lowest temporal frequencies, which is to be expected for bandlimited samples of stimuli — even those with true power law spectra [30,31]. [2] Additionally, two recent studies [23,24], using very different methodologies have also found $f^{-\gamma}$ temporal spectra for sequences of moving images (see below). Is a comparable scaling built into human temporal vision?

3. Correlational structure of human contrast sensitivity

3.1. Background: human sensitivity to temporal scale and modulation

It is commonplace in audio engineering to characterize the frequency response of the human auditory

[2] Power in the lowest frequencies of a time series grows as the length of a series is increased; as the number of periods of data of a particular frequency increases, so does the power detected by the FFT, up to about seven periods of data [30]. Similar limitations apply near the Nyquist limit for noisy data [31]. (Note that both [30,31] discuss optical spatial frequency spectra, but their arguments are general and apply to any discretely sampled bandlimited series of data.)

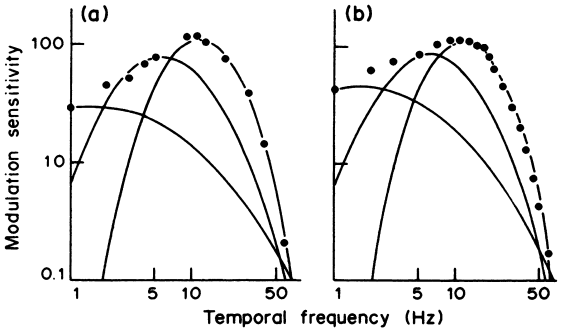

Fig. 2. Temporal contrast sensitivity. Contrast sensitivity functions (CSFs) of human observers vary in shape [37]. Both spatial and temporal CSFs are made up of a number of more narrowly tuned cortical mechanisms (often termed spatial and temporal frequency channels [34–39]). This figure shows the Mandler and Makous [35] three temporal frequency channel model of temporal CSFs (reproduced by permission of Elsevier Science Ltd.). Other models are qualitatively similar. The two frames show how different gains of similar mechanisms can underlie differences in CSFs for two observers (a) and (b).

system and devices designed to stimulate it. Besides corresponding to the wave nature of sound and the human ability to distinguish pure tones, such measures facilitate the application of Fourier's theorem. Similar measures of visual temporal frequency (flicker rate) are of more recent vintage [32,33]. The temporal contrast sensitivity function (Fig. 2) is a plot of 1/contrast needed to detect a sinusoidal modulation of light intensity. In general, humans are most sensitive to moderate temporal frequencies (2–8 Hz) and less sensitive to lower and higher frequencies; for most observers temporal frequency contrast sensitivity (measured for uniform screens or low spatial frequencies) has an inverted U shape — a bandpass filter (see Fig. 2); for high spatial frequencies temporal contrast sensitivity is lowpass. The cuttoff frequency for vision varies with luminance level; humans usually do not detect the 24 Hz flicker of a motion picture in a darkened theater, but do see up to about 80 Hz flicker under very bright illumination levels [33]. Psychophysical studies show that the temporal contrast sensitivity function is not unitary — it consists of several narrower, overlapping, independent mechanisms that together detect temporal modulation [34–36]. One consequence of independence of gain of these tuned mechanisms is that normal human observers vary by as much as a factor of three in contrast sensitivity at any given spatial or temporal frequency [37] (Fig. 2 depicts the mechanisms underlying two observers' contrast sensitivities). Another consequence is that the correlational structure of contrast sensitivity data is a rich source of information about the underlying mechanisms [18,38,39]. Below, we develop and exploit this property of visual perception, in a way that shows an interesting correspondence to the spectra of natural images.

3.2. Characterization of temporal vision correlational structure: methods

We sought to measure the correlational structure of human sensitivity to contrast modulations under dynamic circumstances. This was part of a larger study on contrast perception and more detailed methods are available in Ref. [38]. Briefly, we used 40 observers with a mean age of 26 years and excellent spatial vision (20/10–20/20). All subjects wore their best refractive correction, if applicable. Most were Army aviators or Department of the Army employees.[3] We measured each observer's ability to detect scale specific oscillations in space and time. Such stimuli have the luminance profile

$$L = L_0[1 + M\cos(2\pi f_x x)\cos(2\pi f_t t)] \quad \text{or}$$
$$L = L_0[1 + M\cos(2\pi f_y y)\cos(2\pi f_t t)], \qquad (1)$$

where x and y are spatial dimensions, t is time, f_x and f_y are spatial frequency (cycles/degree of visual angle), f_t is temporal frequency (cycles/s), and L_0 (the space-time averaged luminance) is 100 cd/m^2. These stimuli appear to be flickering gratings of blurry light and dark bars. Such stimuli are used in vision research because they tap scale specific spatial and temporal mechanisms in human vision (they are the psychophysical equivalents of the grating stimuli and pure tones used in Fourier optics and audio engineering). Such mechanisms are orientation specific, with less sensitivity to unoriented

[3] Informed consent was obtained and all experiments were made in accordance with US Army human use protocols.

stimuli, leading us to use oriented stimuli in our psychophysical studies. The overall population of such mechanisms seems to have similar spatial and temporal tuning properties for each orientation, leading us to assume orientation isotropy in perceptual processing for the purposes of this study.[4] These spatiotemporal sine-wave gratings were presented on a linearized CRT display with a large, illuminated screen surrounding the display.[5] We varied contrast (M) and measured temporal contrast detection thresholds for 12 different spatial frequency conditions for each of our 40 observers (about 400 h of data collection). Detection thresholds (M) were measured using interlaced yes/no adaptive staircases to refine an initial estimate generated earlier by the method of adjustment; i.e., a correct detection reduces the contrast of the next presentation, while a missed detection increases the contrast. Such contrast staircases converge on, and oscillate around, the subject's detection threshold; the measured threshold is the average contrast level over the portion of the staircase that oscillates. Spatial frequency of gratings varied from 0.5 to 22.6 cycles/degree (in steps of 0.5 octave) and temporal frequency varied from 0.25 to 32 Hz (in steps of 1 octave).

3.3. Results and analysis

The data were used to calculate the statistical independence of detection thresholds for each pair of temporal frequencies under fixed spatial frequency conditions. A sample temporal frequency correlation matrix is shown in Table 2. This example was created by holding the spatial scale constant at 2 cycles/degree. Each coefficient is the Pearson product correlation between the detection thresholds for two temporal frequency stimuli for our 40 observers. For example, the top left coefficient in Table 2 is the correlation between the detection thresholds of the 40 observers for stimuli flickering at 0.25 and 0.5 Hz. Its relatively high value of 0.58 indicates that subjects who have high sensitivities to 0.25 Hz flicker usually also have high sensitivities to 0.50 Hz flicker, while subjects with low sensitivity to 0.25 Hz generally have low sensitivity to 0.50 Hz flicker as well. However, the correlation between sensitivity to 0.25 and 32 Hz is a meager 0.23, indicating greater statistical independence for the neural mechanisms that detect these more distant temporal frequencies. In general, closer temporal frequencies produce more highly correlated sensitivities than more distant temporal frequencies. This is expected if temporal contrast sensitivity is made up of several statistically independent tuned neural mechanisms (e.g., Fig. 2); stimuli activating a single common mechanisms will have more correlated contrast detection thresholds than those detected by independent mechanisms.[6] A graphical representation (Fig. 3) of this can be created by plotting correlation as a function of temporal frequency separation. If we plot temporal frequency separation logarithmically, all of the correlation coefficients along a given diagonal in Table 2 coincide and can be averaged (each diagonal represents 1 octave of temporal frequency separation). (Although we could have used any spacing for this purpose, a $\log_2$ (octave) spacing

Table 2
Example of a correlation structure derived from temporal contrast sensitivity data[a]

Hz	0.25	0.50	1.0	2.0	4.0	8.0	16.0
0.5	0.578						
1.0	0.682	0.655					
2.0	0.692	0.618	0.859				
4.0	0.368	0.426	0.470	0.567			
8.0	0.283	0.461	0.569	0.702	0.596		
16.0	0.367	0.360	0.509	0.591	0.474	0.718	
32.0	0.231	0.311	0.372	0.493	0.498	0.621	0.677

[a] Table entries are Pearson product correlations of log contrast sensitivities for 2.0 cycle/degree sine-wave spatial frequency gratings flickering at rates of 0.25–32 Hz.

[4] Half of our subjects viewed vertically oriented stimuli, while the other half viewed horizontal stimuli; statistics for both data sets were very similar [38]; we therefore combined the data for statistical power.

[5] Use of an illuminated screen surrounding the CRT display and matched to the luminance and chromaticity of the CRT avoids luminance artifacts at the edge of the display. Such artifacts near the stimuli on the display tend to reduce visibility of the stimuli [34].

[6] This result is not dependent on the bandpass shape of the contrast sensitivity function [38], but does depends on there being multiple contrast detection mechanisms, each covering only a portion of the contrast sensitivity function's temporal frequency range (see Fig. 2).

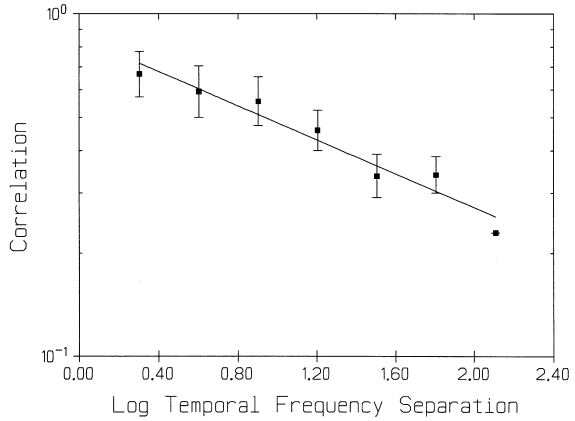

Fig. 3. Example of a correlation function derived from human contrast sensitivity. Our 40 observers show interobserver differences similar to those shown in Fig. 2a and b and these variations result in a pattern of correlations (e.g., Table 2) which would be expected from multichannel models like that illustrated in Fig. 2 (see text for additional discussion). Here, data from Table 2 is collapsed by averaging correlations for constant frequency differences (which are spaced in multiples of log(2); e.g., in octaves). Each point in the graph is the average of one diagonal of Table 2. Error bars show 1 standard error of the mean of the correlation coefficients.

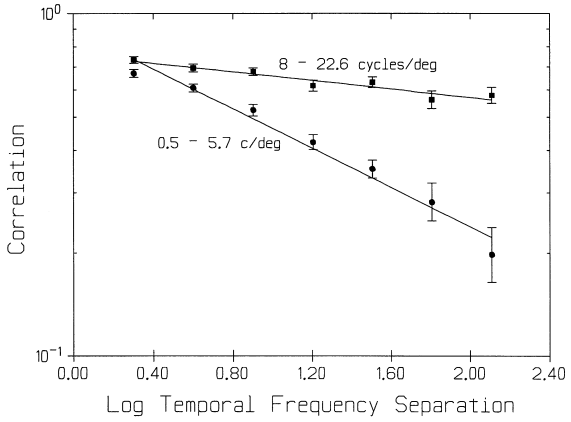

Fig. 4. Effects of spatial frequency on temporal correlation functions. The significant differences ($P < 0.001$) in the scaling properties of low (0.50–5.7 cycles/degree) and high (8–22.6 cycles/degree) spatial frequency sensitive mechanisms may originate in the anatomically and functionally separate parvocellular (small cell, high spatial resolution, low temporal resolution) and magnocellular (large cell, low spatial resolution, high temporal resolution) visual pathways (see Section 4.3).

was chosen because most measurements of visual mechanism bandwidths are given in octaves.) On such a log/log plot, the correlation function is a straight line — a power law relationship between statistical independence and temporal frequency separation. We analyzed these power law temporal contrast sensitivity correlational structures using methodology from the study of fractal systems that have similar correlational structure and $1/f^{\beta}$ spectra [40–42]. Mandelbrot and van Ness [40] showed that fractional Brownian systems with $1/f^{\beta}$ spectra have power law correlational structures (i.e., the correlation between two variables is a power law function of the separation of the scales on which those variables are observed). Bassingthwaighte and Beyer studied similar correlational structures in discrete branching systems [42]. They built on their earlier studies of correlations between blood flows through regions of lung and heart. (Although these are three dimensional systems, the dimensionality is reduced by taking a profile through the system, and assuming tissue isotropy in the subsequent analysis.) They find, similar to Refs. [40,41], a correlation relation which can readily be generalized to continuous spatial or temporal systems:

$$r(d) = 0.5\{|d+1|^{2H} - 2|d|^{2H} + |d-1|^{2H}\}, \qquad (2)$$

where r is the Pearson correlation, d the distance in space, time, or other metric (in our study temporal frequency) and H the Hurst exponent — a measure of long range correlation or trend persistence, ranging from 0 to 1. H is estimated from the slope (S) of the correlation/distance (in our case correlation vs temporal frequency separation) relation (in log/log coordinates) as [16]

$$H \approx 1 + \tfrac{1}{2}S. \qquad (3)$$

The amplitude spectrum[7] exponent associated with such a correlation structure is [16]

$$\gamma = \tfrac{1}{2}(3+S). \qquad (4)$$

Our data showed power law correlational structure for all conditions (Figs. 3, 4 and Table 3), the exponents

[7] Eq. (4) differs from its counterpart in Ref. [16] by a factor of two (the authors refer to amplitude but their mathematics are for power (amplitude squared) spectra). Amplitude (equivalent to stimulus contrast for sinusoidal stimuli) spectra are standard in vision research).

Table 3
Power law model fit to the structure of temporal frequency correlation matrices measured for various spatial frequencies (in cycles/degree)[a]

Spatial frequency	Slope ± S.D.	r^2	Hurst, $(1+\frac{1}{2}S)$	γ, $\frac{1}{2}(3+S)$
0.50	-0.248 ± 0.053	0.812	0.876	1.376
0.70	-0.224 ± 0.042	0.850	0.888	1.388
1.0	-0.243 ± 0.025	0.951	0.879	1.378
1.4	-0.503 ± 0.060	0.934	0.748	1.248
2.0	-0.247 ± 0.026	0.948	0.877	1.377
2.8	-0.306 ± 0.038	0.928	0.847	1.347
4.0	-0.430 ± 0.023	0.986	0.785	1.285
5.7	-0.211 ± 0.011	0.987	0.894	1.394
8.0	-0.074 ± 0.020	0.732	0.963	1.463
11.3	-0.043 ± 0.007	0.864	0.978	1.478
16.0	-0.051 ± 0.012	0.791	0.974	1.474
22.6	-0.094 ± 0.016	0.879	0.953	1.453
0.50–5.7	-0.288 ± 0.022	0.971	0.856	1.356
8–22.6	-0.062 ± 0.009	0.909	0.969	1.469

[a] Slope (S) is in units of log(correlation)/log(temporal frequency separation). The Hurst exponent and amplitude spectra exponent (γ) are computed from the slope (S). The last two rows are computed from data averaged over low and high spatial frequencies (see Fig. 4).

of which are used in Eq. (4) to compute the associated spectral behavior. For low spatial frequency conditions we find γ's of 1.25–1.39 and Hurst exponents of 0.75–0.89, in accord with the high frame-to-frame correlation (0.80–0.86) found for motion pictures [43]. For high spatial frequency conditions (8–22.6 cycles/degree) the correlation data are consistent with a $1/f^{1.47}$ amplitude spectra and the Hurst exponent is a very high 0.97. In fact, a correlation slope of 0 (a γ of 1.5) would be consistent with a single temporal mechanism governing detection at all time scales [38]. Conversely, the higher correlation function slopes found psychophysically for low spatial frequencies agree with evidence for several mechanisms tuned to temporal scale [34–36].

4. Discussion

4.1. Complications

4.1.1. Spatiotemporal inseparability and the spatial characteristics of motion

Recently, Dong and Atick [23] have shown that the spatiotemporal power spectrum of a dynamic image is in general an inseparable function of spatial and temporal frequency.

$$P(f_s, f_t) \propto f_s^{-m-1} F\left(\frac{f_t}{f_s}\right), \tag{5}$$

where m is a measurable property [8] of the ensemble of images in Dong and Atick's [23] video sequences, f_s and f_t are spatial and temporal frequency, respectively, and F is a nontrivial function of their ratio (which under some circumstances can be approximated by a spatiotemporally separable function [21,23]). Although we concur, there are two significant advantages in analyzing temporal spectra in isolation (as done in our study on a pixel-by-pixel basis and in Ref. [24] on a single photoreceptor basis [9]). (a) The slope of the temporal spectra carries information about the character of the image motion. In our five scenes, higher exponents are characteristic of motion which is correlated over all time scales. For example, the highest slopes are for video sequences of people walking or running (γs of 0.87–1.02), in accord with evidence

[8] Dong and Atick [23] measured the spatial spectra of their scenes by digitizing thousands of frames of video and treating each frame as a snapshot, Fourier analyzing each one and averaging over orientation. The resulting snapshot spectra were then averaged and fit by a $1/f^{2.3}$ power spectrum, in agreement with the value inferred from fitting Eq. (5) to the spatiotemporal data.

[9] Van Hateren [24], in a fascinating study, used an artificial detector with a field of view comparable to a single photoreceptor. He mounted it on a helmet and had a subject walk through a natural environment, using head movements to simulate eye movements. The resulting "single photoreceptor" temporal time series (which had a beautifully $1/f$ Fourier spectra) was then used as a stimulus to real photoreceptors (in flies) whose neural responses were recorded.

for long range correlations in studies of human gait [44]. Conversely, the lowest slopes are for the seemingly random motions of small children playing (γ's of 0.61–0.68). For example, film sequence CB — children chasing seeming randomly blowing bubbles — has a γ of 0.61. (b) This space independent formulation may be better suited to the study of temporal information processing in visual cortex. Although the human contrast sensitivity function is spatiotemporally inseparable, analysis of temporal frequency processing in human vision indicates the existence of a number (2–4) of temporal frequency channels, each spatiotemporally separable [34–36] and derived from cortical processing of signals from retinogeniculate neurons with spatiotemporally inseparable contrast sensitivity.

4.1.2. Potential asymmetries in dynamic image spectral assays

As in prior studies of dynamic images [21–24], our spectral analysis of motion pictures makes no distinction of the effects of direction and orientation of motion on the resulting spectra. In one sense, this is appropriate. We are comparing the dynamic spectra of natural images to temporal contrast sensitivity correlational structure that is gathered using flickering stimuli rather than movement. Moreover, although our flicker gratings are oriented, we pooled the data from horizontal and vertical grating stimuli to get better estimates of the correlation coefficients. However, this leaves a significant opportunity for additional study. Although the dynamic scenes that we selected have some vertical motion components (e.g., from rain and drifting soap bubbles and a child climbing on playground equipment), the vast majority of movement in the selected images is in the horizontal direction. In retrospect such a bias seems obvious; we were particularly interested in human movement and walking/running are more often encountered than climbing. Recent studies have looked for orientation asymmetries in spatial vision and found a small but reliable preference for horizontal and vertical stimuli relative to oblique stimuli that may well arise from a similar asymmetry in the orientation of edges in natural and carpentered environments [45–47]. It may be that studies of motion anisotrophies would reveal analogous perceptual phenomena.

4.2. Differences in the correlational structure of spatial and temporal vision

In a previous study we applied a similar analysis to spatial vision [18]. Unlike dynamic images, there is more data and fewer sources of variation in static images and it is possible to estimate the "typical spectra" (the average of a large ensemble of natural images). These studies find ensemble spatial frequency amplitude spectra in the range $f^{-0.9}$–$f^{-1.2}$, with a average over all published studies of $f^{-1.08}$ (1176 total images, see [18] for a review). This raises an interesting question. If humans evolve and develop in an environment with these image statistics, would the correlational structure of spatial contrast sensitivity show a similar built-in sensitivity? For low flicker rates (relatively static conditions), Billock found spatial correlation functions consistent with spectra in the range $f^{-1.09}$–$f^{-1.2}$, mirroring the image analysis [18]. Moreover, the correlational structure of infant spatial frequency contrast sensitivity data [39] does not follow a power law and changes greatly over the first 8 months of life, suggesting that the adult spatial scaling is not built in. Although it would be fascinating to extend this analysis to temporal vision, the paucity of dynamic image spectra makes this unfeasible, and the meaningful motion-linked nature of the temporal spectra (see above), probably makes such an inflexible relationship undesirable. It is interesting however to explore differences in the correlational structure itself. The slope of the temporal correlation functions for every spatial condition (Table 3) are shallower than the slopes of spatial correlation functions for any temporal condition. Following the reasoning of Section 3.3, we could infer that there are fewer independent mechanisms underlying detection of temporal modulation than there are spatial mechanisms underlying detection of spatial modulation. This inference is in agreement with every study of mechanisms underlying spatial and temporal vision (for review see [34]: spatial mechanisms are many in type and relatively narrow in spatial frequency bandwidth (spatial scale),

temporal mechanisms are fewer in type and broader in temporal frequency bandwidth (temporal scale). [10]

4.3. Biological implications of temporal correlational structure

The dichotomous behavior of psychophysical temporal frequency processing at low and high spatial frequencies (Fig. 4) is consistent with Eckert and Buchsbaum's suggestions [22,48] that efficient coding of natural time varying schemes is enhanced by using two functionally and anatomically distinct neural pathways found in early (retina and midbrain) visual pathways: a lower spatial frequency, higher temporal frequency system for detecting stationary (e.g., global flicker) temporal modulation and a higher spatial frequency, lower temporal frequency system for detection of nonstationary temporal modulation (e.g., moving edges). The first system is sometimes called the transient system (because it responds only to the onset and offset of a stimulus) and is mediated primarily by the parasol ganglion cells of the retina, whose signals are relayed to the cortex by the larger (Magno) cells of midbrain. The second system is sometimes called the sustained system (because it responds as long as a stimulus is present) and is mediated primarily by the midget ganglion cells of the retina, whose signals are relayed to the cortex by the smaller (Parvo) cells of the midbrain. [11] Eye movements may keep the components of temporal modulation arising from movement in the frequency range of the mechanisms that subserve them [22,48]. This role for eye movements may explain the mismatch between power law exponents for psychophysical data and moving images. Eye movements tracking image activity would tend to increase the correlation between film frames imaged on the retina, producing retinal image statistics that are more consistent with the psychophysical data. In principle, this could be explored by tracking eye movements induced by an image sequence [49], mapping the fixation point onto the image sequence. A Fourier analysis of the temporal spectra within a moving window centered on the fixation point would simulate the effects on the retinal image of tracking the stimulus. We note that the invariant (power law) form of the amplitude spectra implies additional importance for phase. To the extent that amplitude spectra are interchangeable, phase determines the appearance of events [50] and is a key variable in the interaction of perceptual systems with the environment [17]. Finally, there appears to be a link between the present results, and the power law relationship between physical and perceptual magnitude (Stevens's law). Some psychophysicists suggest that Stevens's law is the result of mapping the output of a logarithmic-like (Fechner–Weber) transduction stage onto a logarithmic representation of the sensory domain [51]. Extrapolated to the frequency domain, such a system would utilize equal neural resources to transduce equal "energy" chunks of a $1/f$ spectrum (e.g., the similar spatial frequency logarithmic bandwidths of cells in visual cortex [5]). Power laws relating input to sensation are likely specific instantiations of the general relationship that we identify here.

[10] It is unclear what implications this has for temporal vision. Color vision relies on only three photoreceptors types (cones) with relatively broad and overlapping absorption spectra to span the 300 nm visible wavelength range, but this suffices to provide us with the wide range of colors that humans experience.

[11] This is a simplification of a rather more sophisticated position. A more lengthy analysis would have to take at least three complications into account: (1) There are at least some Magno cells with sustained responses (e.g. the cells respond to a constant stimulus, not just to a moving or modulated one) [52]. (2) There is good evidence that the *sustained* Magno pathway is involved in the analysis of edge movements [53]. (3) There is evidence that sustained Magno cells become much more transient for high contrast stimuli [52].

Acknowledgements

Supported by NIH grants MH42900 and MH19116. We thank Amber Billock, Rachael Billock, Mark Cannon, Thomas Holyrod, Larry Liebovitch, David Peterzell, Mark Ryder, Brian Tsou and Richard Voss for technical support, helpful comments and other assistance.

References

[1] B.B. Mandelbrot, The Fractal Geometry of Nature, Freeman, New York, 1983.

[2] P. Bak, C. Tang, K. Wiesenfeld, Self-organized criticality: an explanation for 1/*f* noise, Phys. Rev. Lett. 59 (1987) 381–384.
[3] M.F. Shlesinger, Fractal time and 1/*f* noise in complex systems, Ann. NY Acad. Sci. 504 (1987) 214–228.
[4] M. Schroeder, Fractals, Chaos and Power Laws, Freeman, New York, 1991.
[5] D.J. Field, Relations between the statistics of natural images and the response properties of cortical cells, J. Opt. Soc. Am. A 4 (1987) 2379–2394.
[6] G.J. Burton, I.R. Moorhead, Color and spatial structure in natural scenes, Appl. Opt. 26 (1987) 157–170.
[7] D.J. Tolhurst, Y. Tadmor, T. Chou, The amplitude spectra of natural images, Ophthal. Physiol. Opt. 12 (1992) 229–232.
[8] J.H. van Hatteren, Theoretical predictions of spatiotemporal receptive fields, J. Comp. Physiol. A 171 (1992) 157–170.
[9] D.J. Field, Scale invariance and self-similar "wavelet" transforms: an analysis of natural scenes and mammalian visual systems, in: M. Farge, J.C.R. Hunt, J.C. Vassilicos (Eds.), Wavelets, Fractals and Fourier Transforms, Clarendon Press, Oxford, 1993, pp. 151–193.
[10] D.L. Ruderman, W. Bialek, Statistics of natural images: scaling in the woods, Phys. Rev. Lett. 73 (1994) 814–817.
[11] A. van der Schaaf, J.H. van Hateren, Modelling the power spectra of natural images: statistics and information, Vision Res. 36 (1996) 2759–2770.
[12] M.F. Shlesinger, B.J. West, Complex fractal dimension of the bronchial tree, Phys. Rev. Lett. 67 (1991) 2106–2109.
[13] C.K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H.E. Stanley, Long-range correlations in nucleotide sequences, Nature 356 (1992) 168–170.
[14] C.K. Peng, J. Mietus, J.M. Hausdorff, S. Havlin, H.E. Stanley, A.L. Goldberger, Long-range anticorrelations and non-Gaussian behavior of the heartbeat, Phys. Rev. Lett. 70 (1993) 1343–1346.
[15] J.M. Hausdorff, C.K. Peng, Multiscaled randomness: a possible source of 1/*f* noise in biology, Phys. Rev. E 54 (1996) 2154–2157.
[16] J.B. Bassingthwaighte, L.S. Liebovitch, B.J. West, Fractal Physiology, Oxford University Press, New York, 1994.
[17] J.A.S. Kelso, Dynamic Patterns: The Self-Organization of Brain and Behavior, MIT Press, Cambridge, 1995.
[18] V.A. Billock, Neural acclimation to 1/*f* spatial frequency spectra in natural images transduced by the human visual system, Physica D 137 (2000) 379–391.
[19] D.J. Field, N. Brady, Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural images, Vision Res. 37 (1997) 3367–3383.
[20] D.L. Ruderman, Origins of scaling in natural images, Vision Res. 37 (1997) 3385–3398.
[21] M.P. Eckert, G. Buchsbaum, A.B. Watson, Separability of spatiotemporal spectra of image sequences, IEEE Trans. Pattern Anal. Mach. Intell. 14 (1992) 1210–1213.
[22] M.P. Eckert, G. Buchsbaum, Efficient coding of natural time varying images in the early visual system, Phil. Trans. R. Soc. Lond. B 329 (1993) 385–395.
[23] D.W. Dong, J.J. Atick, Statistics of natural time-varying images, Network 6 (1995) 345–358.
[24] J.H. Van Hateren, Processing of natural time series of intensities by the blowfly visual system, Vision Res. 37 (1997) 3407–3416.
[25] P.F. Fougere, On the accuracy of spectrum analysis of red noise processes using maximum entropy and periodogram methods: simulation studies and application to geophysical data, J. Geophys. Res. A 90 (1985) 4355–4366.
[26] G.M. Jenkins, D.G. Watts, Spectral Analysis and its Applications, Holden-Day, San Francisco, 1968.
[27] P.D. Welch, The use of the fast Fourier transform for the estimation of power spectra: a method based on time averaging over short modified periodograms, IEEE Trans. Audio Electroacoustic. A 15 (1967) 70–73.
[28] L.E. Franks, A model for the random video process, Bell Syst. Tech. J. 45 (1966) 609–630.
[29] D.J. Connor, J.O. Limb, Properties of frame difference signals generated by moving images, IEEE Trans. Commun. 22 (1974) 1564–1575.
[30] D.H. Kelly, Spatial frequency, bandwidth and resolution, Appl. Opt. 4 (1965) 435–437.
[31] J.W. Coltman, A.E. Anderson, Noise limitations to resolving power in electronic imaging, Proc. IRE 48 (1960) 858–865.
[32] H. de Lange, Experiments on flicker and some calculations on an electrical analogue of the foveal systems, Physica 18 (1952) 935–950.
[33] D.H. Kelly, Visual responses to time-dependent stimuli. 1. Amplitude sensitivity measurements, J. Opt. Soc. Am. 64 (1961) 422–429.
[34] N.V.S. Graham, Visual Pattern Analyzers, Oxford University Press, Oxford, 1989.
[35] M.B. Mandler, R. Makous, A three channel model of temporal frequency perception, Vision Res. 24 (1984) 1881–1889.
[36] R.F. Hess, R. Snowden, Temporal properties of human visual filters: numbers, shapes and spatial covariation, Vision Res. 32 (1992) 47–59.
[37] A.P. Ginsburg, Spatial filtering and visual form perception, in: K.R. Boff, L. Kaufman, J.P. Thomas (Eds.), Handbook of Perception and Human Performance, Vol. 2, Wiley, New York, 1986, pp. 34.1–34.41.
[38] V.A. Billock, T.H. Harding, Evidence of spatial and temporal channels in the correlational structure of human spatiotemporal contrast sensitivity, J. Physiol. 490 (1996) 509–517.
[39] D.H. Peterzell, J.S. Werner, P.S. Kaplan, Individual differences in contrast sensitivity functions: longitudinal studies of 4-, 6-, and 8-month-old human infants, Vision Res. 35 (1995) 961–979.
[40] B.B. Mandelbrot, J.W. van Ness, Fractional Brownian motions, fractional noises and applications, SIAM Rev. 10 (1968) 422–437.
[41] A.M. Yaglom, Correlation Theory of Stationary and Related Random Functions, Springer, Berlin, 1987.
[42] J.B. Bassingthwaighte, R.P. Beyer, Fractal correlation in heterogeneous systems, Physica D 53 (1991) 71–84.
[43] E.R. Kretzmer, Statistics of television signals, Bell Syst. Tech. J. 31 (1952) 751–763.
[44] J.M. Hausdorff, P.L. Purdon, C.K. Peng, Z. Ladin, J.Y. Woi, A.L. Goldberger, Fractal dynamics of human gait: stability

of long-range correlations, J. Appl. Physiol. 80 (1996) 1448–1457.
[45] E. Switkes, M.J. Meyer, J.A. Sloan, Spatial frequency analysis of the visual environment: anisotropy and the carpentered environment hypothesis, Vision Res. 18 (1978) 1393–1399.
[46] D. Coppola, L. White, D. Fitzpatrick, D. Purves, Unequal representation of cardinal and oblique contours in ferret visual cortex, Proc. Natl. Acad. Sci. USA 95 (1998) 2621–2623.
[47] D. Coppola, H. Purves, A. McCoy, D. Purves, The distribution of oriented contours in the real world, Proc. Natl. Acad. Sci. USA 95 (1998) 4002–4006.
[48] M.P. Eckert, G. Buchsbaum, The significance of eye movements and image acceleration for coding television image sequences, in: A.B. Watson (Ed.), Digital Images and Human Vision, MIT Press, Cambridge, 1993, pp. 89–98.
[49] A.C. Yarbus, Eye Movements and Vision, Plenum Press, New York, 1967.
[50] A.V. Oppenheim, J.S. Lim, The importance of phase in signals, Proc. IEEE 69 (1981) 529–541.
[51] J.C. Baird, E. Noma, Fundamentals of Scaling and Psychophysics, Wiley, New York, 1978.
[52] E.A. Benardete, E. Kaplan, B.W. Knight, Contrast gain control in the primate retina: P cells are not X-like, some M cells are, Visual Neurosci. 8 (1992) 483–486.
[53] K. Nakayama, Biological image processing: a review, Vision Res. 25 (1985) 625–660.